AD-A074 498 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 17/2
VOICE CONFERENCING TECHNOLOGY PROGRAM, (U)
MAR 79 J.W. FORGIE, P.L. WEENE, C.E. FEEHRER F19628-78-C-0002
UNCLASSIFIED ESD-TR-79-78 N/L

1 OF 2
ADA
074498

1 OF 2

ADA 074498

2

8

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

VOICE CONFERENCING TECHNOLOGY PROGRAM

J. W. FORGIE
Group 24

C. E. FEEHRER
P. L. WEENE
Bolt Beranek and Newman Inc.

FINAL REPORT

31 MARCH 1979

LEXINGTON MASSACHUSETTS

CONTENTS

VOICE CONFERENCING TECHNOLOGY PROGRAM

1.0 INTRODUCTION AND SUMMARY

1.1 Introduction

This report has been written at the end of two years of research on voice conferencing technology. The goal of the research has been to recommend and demonstrate the best secure voice conferencing techniques for future defense communication needs. The focus of the work has been on the human factors aspects of conferencing, an area in which little research had been carried out prior to the initiation of this effort. The report has been prepared as a joint effort by Lincoln Laboratory and Bolt Beranek and Newman Inc., who have carried out the human-factors aspects of the research under contract with Lincoln Laboratory.

At the request of the sponsor, the Defense Communications Engineering Center, this report covers work carried out in both years of the program. As a result, some of the material has appeared in previous reports[1,2,3] but is reproduced here to give a complete representation of the work in a single document.

The remainder of this section provides a compact summary of the research program and states the major conclusions and recommendations. Section 2 is an expanded overview of the research that includes a description and discussion in some detail of the conferencing techniques studied in the program. At the end of Section 2, the conclusions and recommendations are restated in a somewhat expanded form. Section 3 gives an overview of the methods and procedures used in the human-factors evaluations of conferencing techniques. Sections 4 through 7 provide detailed descriptions and discussions of the four series of experiments carried out during the program. Appendices give background information on related work and more detail on test scenarios, the simulation facility, and certain conferencing systems.

1.2 Summary

1.2.1 Methodology

In the absence of any established theory applicable to voice conferencing as well as the scant supply of empirical data available in the literature, it was decided that it would be necessary to simulate the various conferencing techniques to be investigated and to evaluate them experimentally. A facility was constructed capable of handling conferences of up to 20 participants. To support the experiments, test scenarios were developed involving group problem solving. Some scenarios provided quantitative measures of productivity. Others provided vehicles for eliciting subjective reactions to the conferencing technique being tested.

Experiments were carried out informally using the researchers themselves as subjects and formally using a group of Lincoln Laboratory employees who volunteered their participation. The informal experiments were used to eliminate clearly unacceptable configurations to save subject time which was a scarce commodity.

1.2.2 Conclusions

In addition to the analysis of the results of the human-factors experiments, analysis has been carried out of the conferencing process itself and of the various system configurations

proposed for evaluation. These activities have all contributed to the conclusions presented in this report. Briefly stated, but with some comment, they are as follows:

1. The requirement to accommodate narrowband users in military conferences necessitates the use of some signal selection technique instead of the conventional signal summation (bridging) technique.

This conclusion follows from two observations:

(a) Summation involves tandem encoding which results in poor quality for narrowband users.

(b) When two or more persons talk at the same time, the result is likely to be unintelligible with present narrowband encoding techniques.

This conclusion, reached soon after the start of the program, caused the research to focus on identifying the best signal selection technique.

2. While summation with analog or wideband PCM signals is superior to any signal selection technique investigated, the use of any of the better signal selection techniques with the same wideband encoding would not result in any significant loss of conferencing capability.

In some test situations using simulated satellite delays and a test scenario focused on collisions (two or more people starting to talk at the same time), selection techniques were given better ratings than summation.

3. Speech quality (voice encoding technique) has a larger effect on subjective judgements of system acceptability than do conferencing protocol (simplex broadcast, speaker/interrupter, etc.) or control techniques (voice control, push-to-talk, etc.).

4. Implementation details such as the operation of speech activity detectors and the procedures for handling collisions in a shared-channel distributed-control system also have a greater effect on acceptability than protocols or control techniques which involve much larger conceptual issues and cost considerations.

This sensitivity to detail suggests that procurement procedures should call for simulation at a sufficient level of detail to check implementation factors.

5. The two most important aspects of conferencing over which a system designer has some control are the ability of the system to handle collisions and the extent to which a speaker may be arbitrarily interrupted. Collision handling appears to be the more important of the two because collisions can be expected to occur far more frequently than situations in which interruption of a speaker is desirable.

6. Experimental results confirm the expectation that large conference sizes do not pose problems for signal selection as they do for summation since noise does not increase with the number of participants.

7. Experiments involving delays of the order of two satellite hops show that such delays have relatively little effect on conferencing performance with the simplex broadcast protocol.

Other protocols such as analog bridge and speaker/interrupter show more effect of delay because a speaker will hear comments and collision fragments with those protocols which are not present with simplex

broadcast. These will arrive at a time when they are not expected and will tend to cause interruptions.

8. Centrally controlled conferencing systems are preferred to those using distributed control of a shared satellite channel, but the difference in ratings is not large, and the distributed-control systems fall well within the acceptable range.

> The differences in ratings are due to the inability of the distributed controllers to handle collisions as effectively as the central controllers. Because of the satellite round-trip delay between the controllers, some speech is lost in a collision before the controllers can detect the collision and take corrective action.

1.2.3 Recommendations

Our recommendations for future secure voice conferencing systems are as follows:

1. Signal selection techniques should be used even though voice-quality considerations for high-level conferences may result in the use of wideband PCM in some systems.

> The small advantage to PCM users of a separate analog bridge for their use would not compensate for the problems a narrowband user would have in connecting to such a bridge.

Voice control should be used with push-to-talk switches gating the voice signals to allow operation in noisy environments. The voice-control algorithm should use a hangover time of the order of 0.4 sec to avoid rapid switching between speakers, an unsatisfactory mode of operation with narrowband encoding. If requirements for "black" controllers so indicate, the push-to-talk switches can be used to control the conference directly with little loss of conferencing performance.

> It should be noted that the use of push-to-talk switches does not imply half-duplex communications. Effective conferencing assumes that full-duplex (4-wire) communications are available and that the listening path remains open when a participant attempts to speak.

2. Centrally controlled conferencing systems should use a simplex broadcast protocol with priority preemption.

> Priority preemption (the ability of a higher priority participant to preempt the conference floor) provides a strong interrupt capability for the higher priority participants. If user needs so indicate, a priority button could be provided to allow for urgent interrupts which are in conflict with the normal priority structure.

3. Recommendations for shared-channel distributed-control conferencing systems depend heavily on the detailed behavior of the communication equipment involved. On the basis of presently available information on equipment characteristics, the recommended technique would be a speaker/interrupter protocol with slow switching of the interrupter channel to inhibit access to that channel until collisions on the speaker channel have been resolved. Collision resolution should use the favored-speaker procedure described in Section 2.7.

An alternative choice which would put less demand on the communication equipment would use a simplex broadcast protocol with an interrupt capability provided by the use of an order-wire channel or by forcing a collision on the shared channel. Further work is indicated to explore other possibilities for distributed-control conferencing. Packet techniques are an example of other communication mechanisms which may yield more satisfactory conferencing than the techniques investigated in this study.

2.0 OVERVIEW

2.1 Statement of the Problem

Future defense communication systems have a requirement to provide a secure voice conferencing capability that is usable by specific high-level command and control personnel as well as ordinary system subscribers and must be capable of expansion into any area where it is needed. Ideally, that conferencing capability would be comparable in voice quality and flexibility to clear voice conferencing such as is provided in commercial telephone systems by performing an analog summation of the signals from the conferees. The requirement for cryptographic security necessitates digitization of the speech signals, and comparable conferencing performance can be obtained if wideband (50 kbps) PCM digitization is used, but cost factors which prohibit general use of wideband digital communications in the near term together with bandwidth limitations in some operational situations result in a requirement for effective conferencing with narrowband speech coding techniques such as Adaptive Predictor Coding (APC) at 9.6 kbps or Linear Predictive Coding (LPC) at 2.4 kbps. There are two serious problems with conferencing by signal summation when narrowband encoding is used. The first is poor voice quality which results from the tandem encoding caused by the necessity to decode the speech at the summing point and reencode the sum for distribution to the listeners. The second is a loss of intelligibility when more than one person speaks at the same time due to the inherent inability of current narrowband techniques to represent the speech of more than one talker at a time. While it may be possible to discover some new narrowband techniques which would tandem satisfactorily and retain some intelligibility with multiple talkers, existing techniques do not do so, and their use with signal summation leads to unacceptable conferencing. We are thus led to explore alternatives to summation as a technique for narrowband conferencing.

The alternatives to signal summation are a large number of possible signal selection techniques. The program for which this document is the final report has been directed toward the examination of these alternative techniques with the goal of identifying, recommending, and demonstrating the most promising technique for use in future defense systems. Factors such as relative cost and complexity have been considered in comparing techniques, but the principal criterion has been the effectiveness of the technique for conferencing. Since there has been relatively little past work on measuring the effectiveness of conferencing techniques, a secondary goal of the program has been the development of appropriate methods for doing so.

There are two basic types of conferencing configurations which are of interest for future systems. The first is centrally controlled conferencing which is schematically represented in Fig. 2-1. In this configuration, each participant has full-duplex (4-wire) communication with a single conference controller. The second type involves distributed control where each participant has his own controller which cooperates with other controllers in sharing the communication medium. Figure 2-2 illustrates distributed controllers sharing a broadcast satellite channel. Central control is an economical configuration for use in terrestrial networks and is used in conventional telephone conferencing. Distributed control can make efficient use of broadcast communications media and has obvious advantages in survivability, but suffers from possible control confusion resulting from communication delay between the controllers and differences in reception conditions at the controller locations. In this investigation, much more attention was directed toward centrally controlled techniques. The only distributed-control techniques investigated in detail have been some that share satellite channels in a fashion similar to one

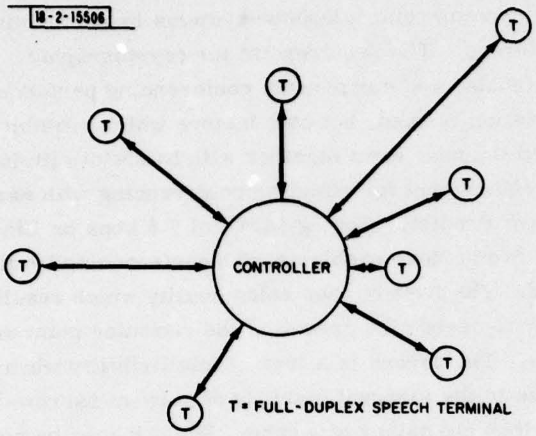


Fig. 2-1. Centrally controlled conferencing configuration.

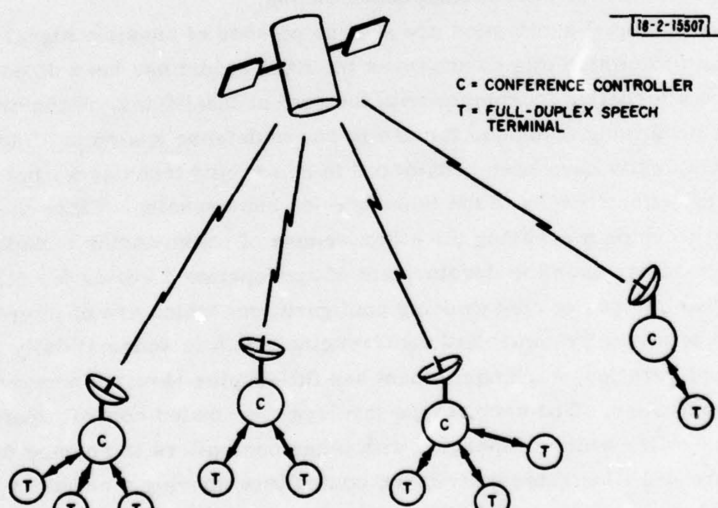


Fig. 2-2. Distributed conference controllers sharing a satellite channel.

proposed for conferencing by World Wide Military Command and Control System (WWMCCS) subscribers. There are many other possibilities for distributed-control conferencing and for mixtures of central and distributed control which it was not possible to study in the scope of this program.

Considerations of interest in evaluating conferencing techniques are the effects of conference size and communication delays such as those introduced by the use of satellites. Other considerations are the effects caused by some of the conferees having different speech encoding equipment or having an extra satellite hop of delay in their communication paths. These considerations have all been taken into account in the analyses and experimental evaluations which have been undertaken, but it has not been possible, nor has it been deemed desirable, to experimentally evaluate all combinations of them.

Future command-level conferences may be expected to make use of dedicated communication facilities and to be augmented by record and graphics conferencing equipment. Such conferences could make use of special equipment for voice conferencing, but since the ordinary subscriber would lack such equipment and could be called upon to participate in a command-level conference, we have assumed that no special equipment for conferencing would be available and have concentrated our attention on conferencing techniques which require only a telephone instrument with push-button dialing and a push-to-talk switch if required by background noise conditions. The push-buttons offer some interesting possibilities for augmenting voice conferencing, and the use of them has been explored in the investigation.

2.2 Conferencing Simulation Facility

Because of the lack of any established theory applicable to voice conferencing as well as the scant supply of empirical data available in the literature (see App. A), we decided at the beginning of the effort that it would be necessary to simulate the various techniques to be investigated and to evaluate them experimentally. To support the simulations, we constructed a conferencing facility capable of handling conferences of up to 20 participants. The facility made use of the Lincoln Laboratory telephone system to allow conference participants to make use of offices at locations with sufficient separation to prevent one hearing another except through the conference phone. Telephone instruments used in the experiments were modified to include dynamic microphones, push-to-talk switches, and tone key pads for signalling. A set of hybrid transformers was used to connect the 2-wire phone lines to the 4-wire equipment of the conference controller which was implemented using an LDVT signal-processing computer to allow the conferencing techniques to be realized in software. Speech input and output to the LDVT were handled by an analog multiplexer-demultiplexer and 12-bit analog-digital-analog conversion at an 8-kHz sampling rate. A large core memory was connected to the LDVT to allow speech samples to be stored for periods of time corresponding to satellite round-trip delays.

To avoid the need for a large number of speech coders, we designed the simulation facility to operate on the PCM samples from the participants until a point at which a speaker had been selected. The PCM samples for that talker were then converted to an analog signal and fed to a back-to-back encoder-decoder pair of the kind called for in the system being simulated. The output of that pair was returned to the controller which then distributed the signal to the conference listeners. With this procedure, it was possible to simulate all centrally controlled configurations with at most four encoder-decoder pairs.

The LDVT conference controller was connected to a PDP-11/45 computer which served as a master control and data collection facility. Information exchanged every 20 msec between the machines allowed the PDP-11/45 to signal the LDVT as to which phones were to be included in any particular experiment and the LDVT to report to the PDP-11/45 as to which phones had signals above a speech activity threshold level, which talker was currently the speaker, etc. The latter information was recorded as a history file on PDP-11 disk storage for later analysis. In some conferencing configurations, the tone keys were used to indicate a participant's desire to speak, etc. In such cases the PDP-11, which was equipped with a special tone key input device, acted as conference controller sending the requisite control signals to the LDVT.

At the end of a formal conference experiment, the subjects were asked to provide rating information which they did by pushing tone keys in response to questions read to them over the conference phones by an experimenter. The PDP-11 made up a file of their responses which was later sent over the ARPANET to a computer at BBN where the data were analyzed. This automated data collection capability was very useful in getting results analyzed in time to affect the planning of succeeding experiments.

Appendix E contains a more detailed description of the conferencing simulation facility as well as a discussion of some of its limitations. The experimental procedures and data collection techniques are discussed in more detail in Sections 3 through 7.

2.3 Procedures Used in Evaluation

Our approach to the evaluation of conferencing techniques has been to gather a group of people and have them participate in conferences using the techniques to be evaluated. The participants are given tasks to carry out during the conference. We call these tasks "test scenarios." A number of different scenarios have been developed over the course of the experimental program. Some are quite simple and require only a few minutes to run. Others are more complex and involve conferences of 20 min. to an hour for completion. Some involve group problem solving and yield quantitative measures of productivity such as solution time and/or quality. Others elicit group discussion of a type similar to that we would expect to occur in a policy-making conference and have no quantitative measure of performance. Some make use of "chairpersons" who have particular roles in the conference similar in character to those normally filled by chairpersons in real conference situations. In our test scenarios, we have taken pains to design the chairperson roles so that the personality of the chairperson is not an important factor in the experiment, i.e., we are trying to measure the effectiveness of the conferencing technique being used, not the effectiveness of individuals as chairpersons.

Because people can readily adapt their behavior to make the best of the situation in which they find themselves, it is often not possible to find differences in quantitative measures of conference performance even when comparing techniques which differ considerably in apparent ease of use. To assess the participants' subjective reactions to the techniques, we asked them to respond to a number of questions about the conferencing situation as well as the scenario being used in the experiment. These responses were elicited at various times during an experiment by asking the subjects to fill in parts of a questionnaire which they were given at the start of the experiment. In early experiments, the questionnaires were collected at the end of the experiment and responses transferred by the experimenter to a computer program for tabulation. In later experiments, the responses were collected directly from the subjects using the tone keying procedure described in Section 2.2 and automatically entered into the computer for tabulation and analysis.

Subjects for formal conferencing experiments were drawn from a pool made up of Lincoln Laboratory employees who volunteered their participation. Experimental sessions were nominally 1 hour in duration and would typically involve from two to as many as seven different conferencing configurations. Conference sizes varied from 4 to 20 participants, but the bulk of the experimentation was carried out with a size of 8 which experience showed to be large enough to exhibit the interesting properties (e.g., several people who wish to speak at the same time) of a large conference but not so large as to make the logistic problems (e.g., getting everyone together at the appointed hour) too burdensome.

Since conducting a formal conferencing experiment is a relatively complex procedure, and our subjects' time was a scarce commodity, we did quite a bit of informal experimentation and evaluation using overselves as participants. In particular, we deemed some configurations as unacceptable on the basis of such informal tests and spared our subjects the frustration of trying to cope with them.

2.4 Signal Summation Versus Signal Selection

As pointed out in Section 2.1, the conventional conferencing technique of signal summation (also called the analog-bridge technique) does not yield satisfactory results when narrowband encoding is used. The alternative to summation is some form of signal selection. There are many possible choices for deciding which talker's speech to select and when to change the selection. We have tried to choose a representative set of the many alternatives for study in this research. All have properties in common which can be contrasted to the conventional summation technique with high-quality speech encoding. The intent of this section is to discuss these general properties before going into consideration of individual techniques.

As anyone knows who has participated in a voice conference, it is not possible to accomplish much when more than one person speaks at the same time. Our analysis of several conferences with signal summation indicated that two or more people spoke at the same time only about 5 percent of the time. Such an observation suggests that a selection technique might be able to produce a good approximation to what a participant would experience in a conference using summation. However, that small percentage of "double talking" time may carry information of importance to the conference. It is important, therefore, to inquire into the content of the double-talking intervals to determine what might be lost (or gained) in going to a signal selection technique.

In our view, the periods of double talking can be usefully divided into the following five categories:

(1) REINFORCEMENT: This category is made up of short exclamations such as "yes," "no," "really?," or non-speech sounds such as chuckles or groans, whose intent is to provide feedback to the speaker. These short sounds do not generally interfere with the intelligibility of the speaker's speech.

(2) COLLISION: This sort of double talking occurs after a pause when two or more conferees attempt to speak at about the same time. The collision may end with all but one talker continuing, or all colliders may cease talking, only to try again after a short pause and perhaps collide again. Intelligibility is likely to be lost in a collision, but the identity of the colliders can often be determined.

(3) OVERLAP: Overlaps occur when a speaker begins talking just before the previous speaker finishes. The new speaker assumes that the previous speaker is finishing from the content and intonation of his speech. Overlaps tend to be short, and intelligibility is likely to be preserved for both speakers.

(4) INTERRUPTION: This category is intended to encompass deliberate attempts to seize the floor by starting to talk before a previous speaker has indicated that he is finished. Intelligibility will become lost if both speakers persist.

(5) NOISE: This category includes coughs, paper rattles, etc., which are sounds not intended to be inputs to the conference. Intelligibility is not likely to be hurt by brief noise.

Of the above categories, reinforcement and overlap are both beneficial to all parties in a conference. Reinforcement provides useful information to the speaker which allows him to adjust to his audience, and overlap allows the conference to proceed more rapidly. An ability to interrupt is obviously viewed as desirable by a would-be interrupter, probably not desirable by the interruptee, and its value to a listener depends on the conference situation. Noises are generally undesirable, and collisions are detrimental to the flow of the conference discussion. It has been our observation that the ability of a conferencing technique to cope with collisions has a major effect on its acceptability.

Summation is always superior to selection with respect to the reinforcement and overlap categories. Some selection techniques allow reinforcement from some one listener to be heard by a speaker. Others require the speaker to pause in order to get any feedback from the listeners. On the other hand, reinforcement while a speaker is talking loses its value when communication delays have satellite hop or larger values because they arrive at the wrong time and tend to disturb rather than reinforce the speaker. As a consequence, people learn to suppress reinforcing utterances when delay is experienced. Selection techniques tend to prevent the use of overlap by either throwing away the overlapping part of the new speaker's utterance or sending it to the previous speaker only. As a consequence, participants either learn to slow down the process of shifting speakers or develop a habit of starting their utterances with throw-away phrases.

Selection techniques are generally superior to summation with respect to ability to cope with noises. In practical conference bridges using summation, it is good practice to use a noise threshold to suppress the buildup of low-level noise which would otherwise grow with the number of participants. This technique, however, does not suppress higher level noises such as coughs which will generally exceed the threshold. With signal selection, such a noise will not be heard unless it is produced by the current speaker or occurs during a period of silence.

Some signal selection techniques are superior to summation in their handling of collision situations. The better selection techniques pick one of the colliders and suppress the others giving listeners clean speech. In many cases, they will be unaware that a collision has occurred. What the colliders hear varies with the technique. Good techniques give an unambiguous indication to a talker as to whether or not he has succeeded in becoming the conference speaker. In this regard, a summation technique with full-duplex (4-wire) communications is a good technique. However, the echo suppression used on long-distance lines causes poor collision handling

in spite of the use of a summation technique because a talker cannot hear the conference while talking and becomes aware of the occurrence of a collision only when he or she pauses.

Some signal selection techniques provide explicit means to facilitate interruptions. Others require a would-be interrupter to wait for the speaker to pause in order to succeed with an interruption. The pros and cons of these techniques will be discussed in ensuing sections when the individual techniques are examined in greater depth. Summation with full-duplex communications performs well on interruption attempts since all parties are made aware of the interruption. As in the case of collision handling, summation with half-duplex communications does not support interruptions gracefully because both speaker and interrupter are unaware of their success or failure in speaking to the conference.

It has been our observation that, overall, the performance of signal summation with high-quality (wideband) encoding and full-duplex communications is superior to any of the signal selection techniques which we have explored. We suspect that the best selection technique would be ranked as superior to summation with half-duplex communication, but we have not conducted experiments with the latter technique since it is not a candidate for future defense system use. With narrowband communications, summation is not acceptable due to voice-quality problems. With intermediate bandwidth waveform encoding such as CVSD, summation may be usable with speech activity detection to prevent noise buildup. We did not examine this case in detail because the requirement to handle narrowband communications forces the choice to some signal selection technique. We did examine a majority voting technique which can accomplish a kind of summation of delta-modulated signals without decoding. Appendix F describes this technique and our experiments with it. We concluded that its application would be limited to a small conference (three or four participants), and that it was therefore not a serious contender for future system use.

2.5 *Control Techniques*

We classify the techniques for controlling a signal-selection conferencing system into three categories. In order of decreasing naturalness and increasing learning difficulty they are:

(1) VOICE-CONTROLLED SELECTION (VC): In this technique, speech activity detectors (SADs) are used to generate control signals on which the conference controller bases it decisions as to which participant should be selected as speaker, etc. A participant need only begin talking to become a candidate for conference speaker.

(2) PUSH-TO-TALK (PPT): In this technique, a conventional push-to-talk, spring-loaded switch on the participant's handset generates a control signal which is sent to the controller. It should be noted that PTT in this context does not imply half-duplex communications as is often the case where push-to-talk equipment is used, e.g., in HF radio communications. In a PTT system, the participant can hear the conference even though his PTT switch is pushed.

(3) CONTROL SIGNAL SELECTION (CSS): In this technique, push buttons on the telephone instrument (tone keys in our implementation) are used to signal the controller as to a participant's desire to talk, etc. The controller maintains a queue of persons waiting for an opportunity to

talk, and when the current speaker finishes talking, it signals the next person in the queue that it is his or her turn to talk. Signals from the controller to the conferee can be either visual or audible. We used an audible signalling technique for our experiments to be consistent with our assumption that the conferee has no special equipment.

VC conferencing depends heavily on proper operation of the SADs. With simple amplitude detectors such as we used in our simulation (see App. E.3.1), VC conferencing can be used only in quiet environments such as offices. While more complex SADs could cope with somewhat noisy surroundings, it is likely that push-to-talk switches would be used where noise is a problem. VC conferencing augmented with push-to-talk switches differs a little from PTT conferencing in that in the VC case the switch merely gates the voice signal whereas in the PTT case the switch directly controls the conference. For example, if a PTT participant holds the switch closed after speaking, he or she will continue to be selected as conference speaker even though no speech is present. This property tends to make PTT conferencing a little more sluggish than VC and more prone to problems caused by inexperienced users.

The potential advantage of PTT over VC with push-to-talk switches lies in the possibilities for secure conferencing without the need to decode and decrypt the speech at the conference controller. In such a situation a "black" conferencing controller could be realized. If either the PTT switch signal or the SAD output can be transmitted to the controller without encryption or on an independently encrypted channel, the encrypted speech can be passed from the sender to the receivers without any intermediate decryption and reencryption at the controller. In such a system, the PTT switch signal, because it changes less rapidly than the SAD output, would require less communication channel capacity, and any timing problems which might result from the transmission of the control signal separately from the speech would be less critical for the PTT case.

In both VC and PTT conferences, speech is lost when participants speak at times when the controller has selected some other participant to be heard by the conference. The CSS technique avoids lost speech by explicitly signalling a participant when it is his or her turn to speak. However, the signalling process slows conference interaction relative to VC or PTT because the signal must have sufficient duration to be detected and the participant takes some time to respond. For example, a simple 8-person word-go-round task (see App. B) which can be carried out in 2.1 min. with a VC system, would be likely to take 2.6 min. with CSS.

In our implementation, the CSS controller maintained a queue of participants who had pushed their "want to talk" buttons and gave them the conference floor on a first-come first-served basis. In our experiments, we observed that while the queue was often empty, it would occasionally grow as big as four or five persons waiting to talk. In such situations, it was likely that the conference discussion would become somewhat less focused than would be the case without a backlog. This defocusing occurs because the order in which speakers are heard is not determined by the current state of the discussion but by the state at some time in the past. For example, if one speaker finishes up by asking a question, the succeeding speaker may very well not be one who has an answer or even cares about the question and will instead start the discussion off in a new direction. Queuing tends to reinforce this behavior pattern by allowing a would-be speaker to sit back and rehearse his speech instead of listening to the conference. Altogether CSS with queueing leads to a "town meeting" style of conference behavior which we

feel is less desirable for problem-solving purposes than the more interactive and focused behavior we observe with VC and PTT techniques.

The most serious difficulty with CSS, and one which leads us to reject the technique, is the difficulty of learning to use the buttons and respond to the signals. While some of our subjects developed reasonable proficiency within a 1-hour training session, others were still having occasional problems after several sessions. Once they had mastered the procedures, they could carry out conferences without difficulty and ended up giving reasonably good ratings to the CSS systems. However, unless a person uses such a system regularly, he is likely to have difficulty and feel frustration that would not be observed with either VC or PTT. Misuse of the buttons or failure to respond to signals tends to cause long pauses in a conference with resulting frustration for all participants. The effort required by a participant in remembering how to use the system is likely to reduce his contribution to the content of the conference. By contrast, VC is very natural to use and requires almost no learning time. PTT requires some practice to master so that the starts and finishes of talkspurts are not clipped, but once learned it does not require much attention during a conference.

The terms VC, PTT, and CSS do not denote three particular conferencing techniques but rather three classes of techniques. A system designer has many options within each class from which to choose. In the course of our research, we have explored a number of these through informal experimentation and analysis and settled on a small set of systems to simulate for use in formal conferencing experiments. In the following two sections, we describe these systems and the rationale for our choices among the options.

2.6 Centrally Controlled Conferencing Systems

Given the configuration of full-duplex (4-wire connections between each participant and the central controller), there are two basic possibilities for signal-selection conferencing. These are:

(1) Simplex Broadcast (SB): The controller selects one participant as "speaker" and broadcasts his or her speech to the others who act as "listeners." The speaker hears nothing.

(2) Speaker/Interrupter (SI): SI is an extension of SB in which one of the listeners can become an "interrupter" and have his speech sent to the speaker. The interrupter continues to hear the speaker. If the interruption is successful (i.e., the speaker stops talking) and the interrupter continues to talk, he will become speaker and the listeners will start hearing his speech. This extension of SB is possible because while a conference speaker is selected the listening channel to the speaker and the talking channels from the listeners are free. The other listeners cannot hear the interrupter because their listening channels are busy with the speaker's speech. SI can be realized with only a slight increase in the complexity of the controller over that required for SB.

There are two system design options which apply to the process of changing speakers. In one case, the selected participant is allowed to continue as speaker until finished. In the other, the system may allow some other participant to preempt the speaker's status from the previously selected participant. The preemption can be based on priority, speaking louder, a timer running

out, pushing a special button, etc. Preemption of a SB conference constitutes an interruption which is much more effective than the kind realized by becoming an interrupter in an SI conference because the change of speakers is forced to occur. Preemption could also apply to the status of interrupter in an SI system, but we do not feel that such a use of preemption would serve any useful purpose and have not simulated such a system. All our experiments which have involved preemption have been with SB systems.

When preemption is used, there is a choice as to who is allowed to preempt and on what basis. Of the many possibilities, we have examined the following SB conditions:

(1) Priority Preemption (PP) with an ordered list of priorities: The participants were arbitrarily ranked in priority and any participant could interrupt all other participants of lower rank. In such a system, the chairperson (if any) would normally be given the highest priority. This technique was explored with both VC and PTT. The use of PP tends to cause listeners to hear more fragments of speech than would be the case without PP. In particular, collisions are likely to be less cleanly handled since small chunks of speech from lower priority talkers may be heard before a late-starting higher priority speaker is finally chosen.

(2) Preemption by the chairperson only: This technique was simulated for VC, PTT, and CSS but was tested in formal experiments only with CSS in an option in which the chairperson had other special keys with which to control the conference. (See App. G for a complete description of the CSS system.)

(3) Preemption on the basis of loudness: A version of a VC/SB system was created which switched speakers whenever the signal level from some other participant exceeded that of the previously selected speaker. If two people talk at the same time, such a system will switch back and forth between them at a rapid rate since the short-term average energy in a speech signal exhibits wide fluctuations at syllabic rates. With wideband waveform coding techniques, such a system will preserve some intelligibility for both talkers, but with narrowband encoding intelligibility is lost in situations in which rapid switching between speakers occurs. We tried to overcome this loss of intelligibility by forcing the system to stick with a newly selected speaker for a time (0.5 sec) long enough to allow a syllable or two to be heard before allowing another change of speaker. This technique helped intelligibility somewhat but not enough to make preemption on loudness a workable technique for narrowband use. It was examined in one early experimental session and eliminated from further consideration.

It has been our observation that for both VC and PTT techniques SB systems are almost always superior to SI systems because SB handles collisions in a much more satisfactory fashion, and collisions happen much more frequently than do needs to interrupt a speaker who refuses to pause and give others a chance to talk. With SB, when a would-be talker hears someone else speaking he knows that he is not the selected speaker and that he should stop talking and wait for another opportunity. With SI, however, hearing someone else is not a good indication

of failure to get the conference floor since the voice being heard may be that of an interrupter. In such a situation, all colliders may back down and try again, perhaps to collide again. Alternatively, a talker may persist, thinking that he has the floor, and his role in the conference may then suffer because the utterance he produced was not in fact heard but he thought it was and did not act to repeat it at a later opportunity. In the case where the colliders back down, listeners hear fragments of speech from one or more of the colliders. With SB, the selected speaker is unaware that a collision is occurring and proceeds to finish his utterance without difficulty. Listeners hear fewer fragments of speech, and the role of listener is more pleasant.

In conferencing experiments, we record the speaker and interrupter signals on separate tracks on magnetic tape and can therefore listen to what is heard by listeners as well as speakers in a conference. One observation to be made from listening to the interrupter track is that almost nothing recorded there could be construed as intended to be heard only by the conference speaker. Most of the material is fragmentary, but there are occasional complete utterances which were clearly intended to be directed to the conference as a whole. In one experimental session involving the "car pool" problem-solving scenario (see App. B), the answer to a question was given on the interrupter channel and the talker thought it had been heard by the conference. A period of several minutes went by before it was discovered that the information was missing, and the answer was repeated. We feel that this confusing property of SI systems is undesirable.

In comparing subjective judgments of SI and SB systems, it should be noted that if during the course of an experiment there were very few collisions or attempted interruptions, the systems would be indistinguishable since there would be little or no speech on the interrupter channel. The consensus scenario used in comparing centrally controlled systems could result in more or fewer collisions and interruptions depending upon the degree of involvement which the particular discussion topic engendered. Some discussions were relatively heated and produced as many as 30 or 40 collision events in a 5-min. period. Others were quieter and produced only three or four collisions. We could expect SI systems to be less well liked when collisions were more frequent, and we found this to be generally true. In no case did we find a significant preference for SI over SB.

2.7 Distributed-Control Conferencing Systems

In this program, we have investigated a class of advanced conferencing techniques similar to one proposed for use in the World Wide Military Command and Control System. These techniques all make use of shared broadcast satellite communication channels. Transmission uses spread-spectrum techniques for jam resistance, and all speech is encrypted for security. We use the term SCDC (Shared Channel with Distributed Control) to refer to systems of this type. A simple SCDC conference would involve a single satellite with a number of earth stations each of which has a conference controller which could support a number of participants who are viewed as being local to that controller. The controller would use a central voice control protocol to select among its local participants and a distributed protocol to interact with the other controllers to decide which participant can use the satellite channel. A more general configuration might involve multiple satellites with some controllers serving as linkers or gateways between the satellites.

Because the complexity of the simulations required to model the shared-channel behavior pushes the capacity of our conferencing simulation facility, we have had to reduce the number of participants which can be handled in comparison with the centrally controlled systems. We

have also reduced the simulation load by assuming that each earth-station controller serves a single conference participant. This simplification allows us to have a larger number of earth stations in the simulation and therefore focuses attention on the sharing of the satellite communication channels which is the distinctive feature of SCDC conferencing.

There are three basic protocols of interest for SCDC conferencing. They all make use of voice control and are:

(1) Simplex Broadcast (SB): This protocol is very similar to the simplex broadcast protocol in centrally controlled conferencing. A talker is selected as speaker and his speech is broadcast to all others. The speaker hears nothing. SB requires only one satellite channel and one set of transmit/receive equipment at each earth station.

(2) Speaker/Interrupter (SI): Again, this protocol is very similar to centrally controlled SI with the exception that the time available to a would-be interrupter is reduced relative to the centrally controlled case by the time required to achieve synchronization of the interrupter channel. The speaker can hear a single interrupter; but if more than one participant attempts to become an interrupter, the speaker will hear either noise or nothing at all depending upon the behavior of the crypto equipment. SI requires two satellite channels but only one set of transmit/receive equipment.

(3) Broadcast Interrupter (BI): This protocol requires two satellite channels and two sets of transmit/receive equipment at each controller as well as an extra speech decoder at each participant's site. It allows the speaker to hear the interrupter, the interrupter to hear the speaker, and listeners to hear both by summing the speaker and interrupter signals.

SCDC conferencing differs from centrally controlled conferencing in two important respects. The first is the delay which results from the round-trip time to the geostationary satellite (270 msec) plus the encryption preamble time which depends upon the speech encoding rate and crypto technique. We have used preamble times from 24 msec to 1.07 sec in our simulations. Speech is stored at the sending controller during the preamble transmission time. The second difference comes from the inability of the distributed controllers to make ideal decisions about which talker to select. SCDC conference control depends on sensing the absence of signals on the channel to allow a new talker to use the channel. Because of the communication delay, there is a window of 270 msec after a controller has started transmission before the other controllers become aware of the transmission. During this time, one or more of them may also decide to transmit. In that event, interference will occur and all ground stations will receive noisy signals that will cause one of two possible events. If the channel collision occurs during the time that the first controller is sending the crypto preamble, crypto synchronization will not occur at the receiver and the listener will hear nothing. We call this case a major collision. If the collision occurs after the first talker's preamble has been successfully transmitted, crypto synchronization will be achieved at the receivers, but the listeners will hear a burst of noise which will continue until either the colliding controller stops transmitting or the crypto device decides that it has lost synchronization and shuts off the output speech. We call this second case a minor collision, since the duration of the noise burst can be kept short by proper

controller action. With proper controller action, minor collisions can occur only in situations in which the preamble time is less than the satellite round-trip time or in the special case when they occur on the interrupter channel of an SI system. In the latter case, collisions cannot be detected because the colliding interrupters do not have extra receivers with which to listen to the interrupter channel. Their available receivers are set to listen to the speaker channel.

There are many possibilities for controller algorithms to handle collisions. We have investigated a number of these and chosen some for experimental evaluation. They all depend upon the ability of the controller to sense the presence of a carrier signal on the channel as well as to determine the successful acquisition of synchronization by the crypto equipment. They all use this information in the same way to detect collisions. The differences lie in the actions taken after a collision is detected.

In the usual case, a controller is permitted to start transmitting on a channel only when it finds the channel to be free, i.e., a carrier is not being received. Upon starting to transmit the crypto preamble, the controller starts watching the channel receiver for the detection of a carrier signal. If the controller's transmission is successful, i.e., no collisions occur, it will first detect the carrier one round-trip time after the transmission started. Actually, a small additional time is required for the receiving equipment to reliably detect the presence of the carrier, but this time is assumed to be short compared with the round-trip time and has been neglected in our simulations. If the controller gets carrier detection sooner than one round-trip time after starting transmission, it knows that it is in collision with some other controller that started transmitting earlier. Its response to early carrier detection is to immediately cease transmission to minimize the time during which the channel will be unusable due to the collision. With this algorithm, the period during which the channel is actually in collision will be at most one round-trip time. However, the controller that started transmitting first and detected the carrier at the expected time will not become aware of the collision until it fails to achieve crypto synch after an additional preamble time has elapsed. In the case where the preamble time is shorter than the round-trip time, it is possible that the first controller may have transmitted a complete preamble before any other colliders started transmitting. This is the minor collision case. Crypto synch will be achieved and listeners will hear a burst of noise of a duration equal at most to one round-trip time minus the preamble time. In the case where the preamble time is equal to or longer than the round-trip time, all collisions will be major. Crypto synch will fail and the controllers, if they still have speech signals to send, will enter some algorithm for trying again to use the channel.

The differences in the control algorithms lie in the procedure used in trying again. Of the many possibilities, we have examined four in some detail. They are:

(1) Favored Speaker – Version 1 (FS-1): This algorithm allows only the first collider (favored speaker) to try again as soon as the channel becomes free. The controllers decide who is first according to the time at which they detected the collision as described above. If two talkers started speaking at exactly the same time, say within a few milliseconds, their controllers will conclude that each is the favored speaker and will attempt to use the channel again, causing another collision, etc. The probability of such an almost simultaneous start is much smaller than that of two or more starting within a round-trip time, and we feel that it can be ignored

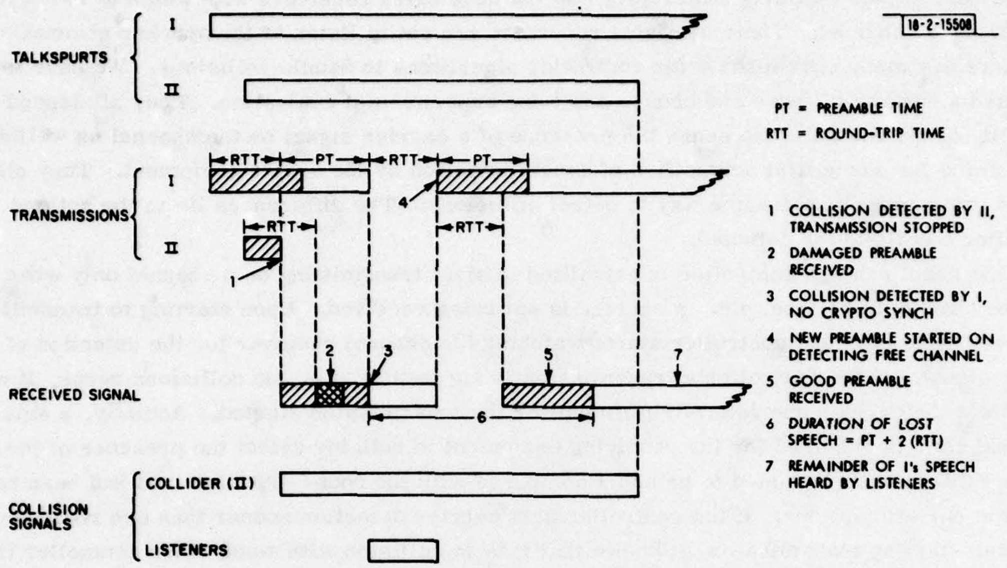


Fig. 2-3. SCDC collision-handling algorithm FS-1. Favored speaker (I) proceeds when channel becomes free.

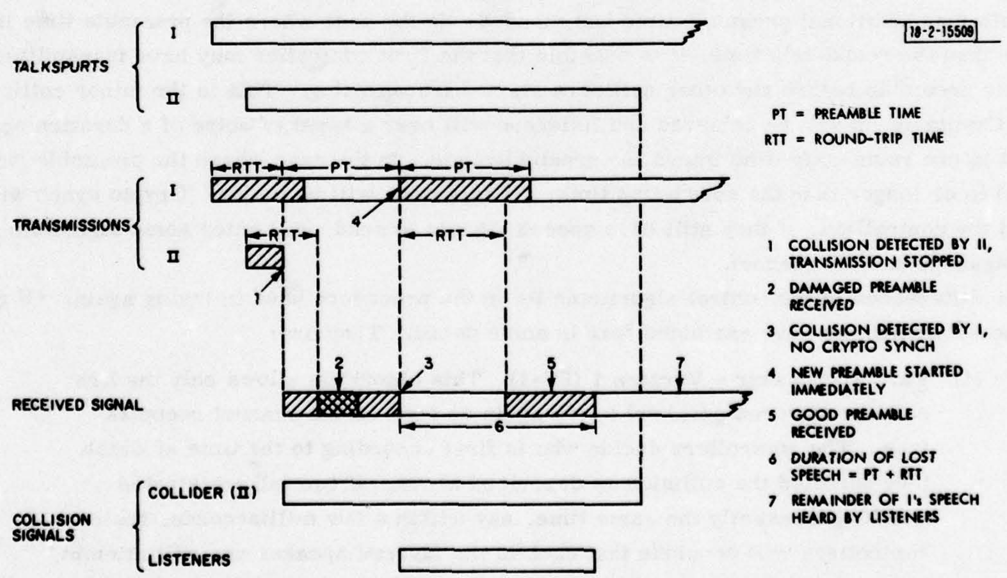


Fig. 2-4. SCDC collision-handling algorithm FS-2. Favored speaker (I) proceeds immediately.

in practice. We are not aware of the occurrence of such an event in any of our experiments. With FS-1, the colliding controllers that have determined themselves to be late colliders are not allowed to transmit again until their talkers become silent and then start speaking again. The action of this algorithm is indicated schematically in Fig. 2-3 for the case of a two-talker collision. The time lost on the channel due to a collision handled by this technique is one preamble time plus two round-trip times. Listeners will miss that much of the talkspurt of the favored speaker. Otherwise, the behavior is very much like that observed for centrally controlled conferences with equivalent overall delay.

(2) Favored Speaker – Version 2 (FS-2): FS-2 is similar to FS-1 in allowing only the first collider to retry, but it does not wait for the channel to become free. Instead, FS-2 lets the first collider begin retransmitting the crypto preamble as soon as failure to achieve crypto synch is detected. The action of this algorithm is represented schematically in Fig. 2-4. FS-2 represents conceptual advantages over FS-1 since it reduces speech loss to the preamble time plus one round-trip time (an improvement of 270 msec), and it prevents any new colliders from contending for the channel by keeping the carrier on until the favored speaker finishes his talkspurt. However, as one might expect, the advantage perceived by our experimental subjects is not significant since the extra 270 msec of speech heard with FS-2 does not contain much useful information, and there is still an awareness that speech is missing.

(3) Free For All (FFA): FFA allows all colliders to try again as soon as the channel becomes free. This algorithm represents an almost uncontrolled use of the channel. Communication is lost until all but one of the colliding talkers becomes silent. FFA is simpler than the FS algorithms and can be used in situations such as might occur in the presence of jamming where the carrier detection required in FS may not be reliable.

(4) Random Suppression (RS): RS allows all colliders to try again but with some probability less than one. The intent of this algorithm is to improve on the FFA algorithm by increasing the probability that a retry will be successful. Communication is lost until either all but one of the colliders becomes silent or random choice results in only one controller trying again. With the probability of retry set at one half, RS did not offer any noticeable improvement over FFA.

The technique used in the FS-2 algorithm of transmitting even though the channel is not observed to be free is also used in all algorithms in the case where a controller has been transmitting, its talker has gone into silence, transmission has stopped but a round-trip time has not yet elapsed, and the talker starts speaking again. In this case, the controller knows that the

channel is busy with its previous transmission, and that it may transmit again without risk to the material being received. It is important for a controller to allow transmission in this case because otherwise talkspurts by the same speaker which succeed others by more than a hangover time but less than that plus a round-trip time would be clipped or lost. Such a sequence of talk-spurts is likely to occur in situations such as spelling a name or saying a string of digits slowly and precisely. If the controller waits for the channel to be free, every other word will be lost in such a case even though there is no contention for the channel. Behavior of this kind would make a conferencing system unacceptable for use in military conferences where such speech patterns may be expected to occur frequently.

An important adjunct to the algorithm for controlling collisions is the provision of some kind of signal to the conference participant to indicate that the channel is in collision. Such a signal might be visual or audible. We have experimented only with the use of an 800-Hz tone for such signalling purposes and have not carried out any studies to optimize the signal as to quality or intensity. The signal options evaluated were the following:

(1) No beep: In this condition, participants hear nothing when the channel is in the major collision state. A participant who tries to become the conference speaker has no indication of success. If he hears another participant he knows he has failed, but hearing no one does not indicate success as it does for a centrally controlled SB system, for example.

(2) Beep to disallowed colliders: In the FS algorithms above, the favored speaker hears no beep. The other colliders hear a beep that persists until they become silent. The beep duration corresponds to the period in which they are being denied access to the channel. In the FFA and RS algorithms, all colliders hear a beep that persists until the channel becomes free. In a repetitive collision situation the channel becomes free periodically between collisions, and a talker who keeps speaking will hear a periodic beep which will continue until he succeeds in getting the channel or gives up. In this option, passive listeners hear no beep and are unaware that collisions are occurring. In the case of a BI protocol system, collider beeps are heard for collisions on either of the two channels. In an SI system, a controller attempting to use the interrupter channel has its only receiver set to the speaker channel and therefore cannot listen to the interrupter channel to determine if it is already in use or to carry out any collision detection or resolution algorithm. Consequently, no beeps can be generated for collisions on the interrupter channel in such a system. The conference speaker will hear a noise in the event that a collision occurs on the interrupter channel, but the colliders will be unaware that a collision is taking place.

(3) Beep to listeners as well as colliders: Colliders hear beeps as in Option 2, but in addition passive listeners hear short beeps which start as soon as the listener's controllers detect failure to acquire crypto synch and last until either the channel becomes free or crypto synch is acquired. The listener beeps were suggested by some of

our experiment subjects who after experiencing Option 1 and 2 felt that the additional beeps would provide useful information. Experimental evaluation confirmed the desirability of these beeps. It should be noted that listener beeps would not be feasible in situations such as jamming where carrier detection might not be reliable. Listener beeps were provided only for collisions on the primary channel in BI and SI systems.

The SCDC voice control technique is more demanding than centrally controlled techniques with respect to performance of speech activity detectors. In a centrally controlled system, when a talker has been selected as speaker, the controller will leave him in the selected state even though his voice energy drops below threshold, and he will remain selected until some other talker goes above threshold. A speaker who tends to trail off in amplitude toward the end of his utterances will be heard to completion unless some other participant starts talking before he finishes. With the SCDC technique, however, it is necessary for the speaker to stop transmitting when he drops below threshold in order to allow some other participant to become speaker. If he trails off in amplitude and falls below threshold before fininshing his utterance, he will be cut off even though no other participant wants to speak. To prevent such cutoff it is necessary to use a relatively low SAD threshold at the end of a talkspurt which makes the system less robust with respect to acoustical background noise. The SADs in our simulations have different starting and ending thresholds which allow the starting threshold to be set relatively high to suppress noise and the ending threshold to be set low to stay with a speaker who trails off. This technique works adequately well under quiet conditions, but we have found it desirable to further minimize problems with background noise by using push-to-talk switches to gate the microphone signals. We feel that the use of such switches would be good practice in field application of all VC conferencing systems, and that their use is particulary desirable in SCDC systems where staying above threshold due to noise can tie up the channel, and priority preemption is not possible to recover conference control.

2.8 Conference Augmentation

In addition to providing voice communication, a conferencing system can provide other aids to group problem solving. Future command-level conferences are expected to be supported with equipment which can distribute typewritten and graphical material for the use of conference participants. Without requiring special equipment at the subscribers location, it is possible to use the push button available on his telephone to send signals which a conference controller could use for a variety of functions to aid the work of the conference. Some are related to the operation of the conference controller. These include signalling departure from and reentry to an ongoing conference, and requests to change priority or move the location of a chairperson. Others can be used to speed the flow of the conference discussion. These include vote taking and indicating the extent of agreement or disagreement with a position taken by some speaker.

The value of these aids depends upon the detailed nature of the task being performed by a conference and is not readily assessed in a laboratory environment. We have not attempted to conduct formal experiments to evaluate conference augmentation, but we have accumulated considerable experience with vote-taking procedures in connection with gathering our subjects' responses to questions regarding the performance of the conferencing system they have been using. At the completion of each scenario a special system is loaded into the simulation facility,

and an experimenter seated at a computer display console reads a few key words of each question on a questionnaire which the subjects were given before the experimental session started. As each question is indicated by the experimenter, the subjects each push a tone key corresponding to the response they wish to make to the question. The display indicates to the experimenter the response that each subject has made to the question. When all subjects have responded, the experimenter moves to the next question. The procedure is partially automated with the computer providing phrases to the experimenter and moving to the next question when all have responded. It could be fully automated with the computer asking the questions using stored prerecorded or synthetic speech, but we have not done so. This procedure works smoothly and rapidly with very little training required on the part of either the subjects or the experimenter. The same technique could be used to take votes or preference judgments during conferences with a considerable gain in speed over the usual procedure of having a chairperson or secretary poll each participant in turn. However, this type of voting requires some special equipment to display the results of the vote and some special knowledge on the part of the person who operates the equipment.

We feel that augmented conferencing is likely to require the services of a conference operator who would have access to and training in the use of the special equipment required. Such an operator could aid in setting up conference calls, but such aid should not be required in future communication systems which should automatically handle the setting up of connections among the participants and the controller(s). There is no need for an operator handling augmentation to be a conference participant. He or she could have separate voice communication with the conference chairperson or secretary so that requests for service could be handled, but the conference content could remain private and secure.

2.9 Formal Experiments: Centrally Controlled Techniques

Most of the formal experiments concerning centrally controlled conferencing techniques were carried out in two series of experiments. The first (Phase I) addressed issues of conference size and overall delay and compared the VC/SI technique with a signal-summation technique, the traditional analog bridge. The second series (Phase II) involved only signal selection techniques and compared CSS, PTT, and VC using both SB and SI protocols with a variety of options. Phase II also examined the effects of extra communication delay to a subset of the conferees and the effects caused by tandem speech encoding when a subset of the conferees has different encoder/decoder equipment. At the end of Phase II, a pair of experiments were carried out to compare the centrally controlled techniques with SCDC distributed-control techniques. Additional experiments examined the behavior of large conferences (20 participants) using a different scenario. In this section, we state the condition for these experiments and briefly summarize the results. More detail on the experimental procedures and analysis of results can be found in Sections 4 and 5.

In Phase I, the "car pool" resource allocation scenario (see Sec.3 and App.B for descriptions of scenarios) was used. This scenario was a problem-solving task where all participants had an equal role. Problem solution time was on the order of 20 min. There was no chairperson or other special role in the conference. In order to be able to compare signal selection with signal summation, PCM encoding was used in all experiments. Subjects were asked to rank the systems as to relative difficulty of use.

In Phase II, a group discussion scenario was used. One participant was chosen as chairperson and given the task of getting the group to reach a consensus on the solution to a hypothetical problem. Discussion time was limited to 7 min., and one of the chairperson's tasks was to bring the discussion to a halt on a signal from the experimenter and then to summarize the position of the group and poll them to determine to what extent a consensus had been reached. At the end of the experiment, the subjects responded to a questionnaire (Sec. 5.2.3) by pushing their tone keys. The chairperson was asked some additional questions relative to his or her special role in the conference. The results yield an estimate of the subjects' perception regarding the relative acceptability of the systems. In this series, 32-kbps CVSD speech encoding was used except when tandem encoding was a system variable. All conferences involved eight participants, and a communication delay of 0.5 sec for all participants was simulated except when extra delay was a system variable.

In the large conference experiments, the Telewar scenario was used. Telewar is a highly structured military route-finding scenario with a chairperson, planners, and staff people. Rating information was gathered using the same questionnaire and technique used in the other Phase II experiments. Speech encoding was 32-kbps CVSD and a delay of 0.5 sec was simulated.

2.9.1 Effects of Conference Size

In Phase I, experiments were run with conferences of 4, 8, and 12 participants. As might be expected, conferences increased in apparent difficulty with size both because communication became somewhat more difficult and because the conference task became more difficult since the number of commuters to be dealt with in the car-pool allocation increased as the number of conferees increased. Subjects reported that they adopted a more disciplined or formal style to deal with the increased likelihood of collisions in the larger conferences. In many cases, some participant would for a time assume a role like that of a chairperson to regulate the flow of the conference.

Increasing probability of collision is the principal source of increasing communication difficulty as conference size increases. The chance that a collision will occur depends upon the structure (or lack of it) in a conference. For example, there are very few collisions in the Telewar scenario because even though there may be 20 participants in the conference and all will play some role over a period of a half hour or more, at any one time only 3 or 4 are actively engaged in exchanging information. The others are merely listening and waiting for a point in the problem solution at which the information they possess will be needed. We believe this behavior is characteristic of large conferences which have to be structured to be productive even in face-to-face situations.

We have drawn two conclusions from our experiments with various size conferences. These are:

(a) There is no real limit to the size of conferences which can be handled with signal selection techniques.

(b) Conference sizes of the order of 8 to 10 participants are sufficiently large for the purpose of testing conferencing systems. Larger groups do not produce significantly higher probabilities of collision or introduce other problems to challenge the conferencing technique.

2.9.2 Effects of Delay

In Phase I, experiments were run to assess the effects of delay on conferencing. Behavior with no delay was contrasted with behavior when communication delays of 0.5 sec were simulated for all participants. Such a delay corresponds roughly to that which would be experienced if all participants were one satellite hop away from the conference controller. On first exposure to the delay situation, the subjects had more trouble with collisions than had been the case without delay. For a time they exhibited a tendency to repeat themselves in an effort to make sure that they had been heard, but they soon adapted to the delay and the conference proceeded at a pace comparable with that observed without delay.

Delay has the effect of prolonging any problems that may result from collision situations since more time must elapse before the colliders discover that a collision has occurred. In the case of an analog bridge or SI protocol system which allows reinforcing sounds to be heard by the speaker, delay acts to negate the benefit of such reinforcements since they arrive at a time appreciably later than the point in the speaker's utterance which triggered them. Delay of 0.5 sec or more is likely to cause the speaker to treat an attempt at reinforcement as a desire to interrupt because the sound he hears does not come when he expects it, but at a later time when he is already into the next phrase of his utterance. People rapidly become aware of this aspect of delay and change their behavior to suppress reinforcements except when explicitly prompted by the speaker who then waits for the requested feedback.

In the Phase I experiments, conferencing with delay was perceived as being more difficult than conferencing without delay, though we observed no statistically significant differences in group performance. In Phase II, we chose to do all experiments with delay on the grounds that the increased difficulty might intensify any differences caused by other factors being investigated. As a result, our subjects became quite accustomed to the effects of delay and several commented that they had become completely unaware that delay was present and noticed it only on occasions when another participant might speak loudly in a nearby office and be heard directly before being heard on the conference handset.

In some later experiments using SCDC systems with long preamble times, subjects experienced overall delays of about 1.3 sec. The subjects found these long delays to be annoying and gave poor ratings to the systems which had them, but task performance was not affected in a major way, and we feel that systems with such long delays would be acceptable if communication conditions require them.

Another effect of delay on conferencing occurs when come subset of the conferees experiences extra communication delay relative to the others. We explored this situation in several experiments in Phase II by adding an additional 0.5-sec delay to the speech of one talker. To maximize the effect, we selected the conference chairperson as the disadvantaged speaker. The results do not show a large effect for any of the tested cases, but as might be expected the most noticeable difference occurred with a VC/SB system. With voice control, a participant with extra delay will arrive late and fail to get the floor in all situations where one or more participants try to talk at the same time. Again, as could be expected, the results showed that giving the delayed participant an ability to preempt the floor could compensate, at least in part, for the disadvantage of extra delay. It should be noted that the impact of extra delay on a particular experiment depended upon how often collisions occurred between the disadvantaged participant and others. If no such collisions occurred, there would be no awareness of the extra delay and any differences in ratings would be due to other effects or random variability in the ratings.

Our conclusions with respect to delay effects are the following:

(a) Overall delays of the order of 0.5 sec have very little effect on the pace of a voice conference using signal-selection techniques.

(b) Extra delay for a subset of participants puts those participants at some disadvantage in competition for the speaker's position in a conference. This disadvantage is often insignificant, and it can be compensated for to some degree by giving the delayed participants a higher priority in systems where preemptive priority is used.

2.9.3 Effects of Speech-Encoding Technique

In the course of this program, conferences have been carried out with four different speech-encoding techniques. There were:

(a) LPC at 2.4 kbps
(b) APC at 8.0 kbps
(c) CVSD at 16 and 32 kbps
(d) PCM at 96 kbps.

The Phase I experiments were done using PCM to allow comparison with the analog bridge. Phase II used 32-kbps CVSD except for experiments involving tandem encoding in which case one participant had an APC encoder at 8 kbps. In the tandem experiments, a CVSD listener heard the speech of all CVSD speakers through a single CVSD encode-decode process. However, when the APC participant spoke, the CVSD listener heard speech that had been passed through an APC encode-decode followed by a CVSD encode-decode. The APC participant heard the reverse tandem when a CVSD speaker spoke. If there had been other APC speakers in the experiments, they would have heard each other through a single APC encode-decode process.

LPC encoding was used in early informal experiments that led to the conclusion that rapid voice-controlled switching between speakers was unsatisfactory for narrowband encoding. Consequently, all other experiments used a slow switching technique requiring a silent interval of 0.4 sec before switching away from a speaker.

Results from the tandem experiments showed that speech quality was the strongest factor in determining the rating which subjects would give to a conferencing technique. It had a much greater effect than conference size, delay, or any protocol variables. The only stronger effect we observed was improper operation of the speech activity detectors in a voice-control system which could (and did on a couple of occasions) make systems unusable. Even with the CVSD-APC tandem the subjects had no difficulty in carrying out the conference task, and conferencing with such a tandem may be considered to be acceptable though clearly less desirable than conferencing using a uniform encoding technique.

We believe that our results regarding the effects of encoding techniques are conservative because the effective speech quality experienced by participants in our experiments was less good at all rates than one would expect to experience in a true 4-wire digital communication system using the same encoding techniques. The reduction in quality came about because of the use of the telephone system in our simulation facility which introduced noise and distortion at the input to the speech encoders. Further quality loss resulted from the imperfect operation of the hybrid transformer which matched the 2-wire phone system to the 4-wire simulator and from some aliasing that occurred in the analog-to-digital conversion process.

2.9.4 Effects of Control Techniques

The Phase II experiments showed that subjects preferred VC over CSS and PTT systems but not by a large margin. As stated earlier, we reject CSS as a viable candidate for future systems because of the difficulty in learning to use the technique, not because it does not work well once learned. VC and PTT or a mixture of the two techniques can be rated as almost equally viable candidates with the choice being left to considerations other than human factors, such as background noise or the potential for "black" conference controllers.

2.9.5 Comparison of Conferencing Protocols

The Phase II experiments provided many opportunities to compare the effectiveness of the SB and SI protocols under a wide range of conditions. SB systems were tested with and without priority preemption. SI systems were tested with all participants allowed to be interrupters and with only the chairperson so allowed. VC/SB systems with and without priority preemption were preferred over all others, but the margin of preference was not great. There was much overlap in judgments, and the variance in mean scores is such that many systems must be considered as essentially equal in acceptability. (See Sec. 5.3 for a more complete presentation of the experimental results.)

Among the SB systems there is not a clear preference either for or against the priority preemption option. As might be expected, the participants who had higher priority liked the priority preemption systems better than those who had lower priority. Listeners are apt to prefer a system without preemption because they will hear fewer fragments of speech with such systems. As mentioned above, priority preemption can partially compensate for the disadvantage of extra delays. Priority preemption allows easy interruption by high-priority talkers, but that advantage did not have much effect on the ratings even though the scenario required the chairperson to interrupt the conference on occasion. In most instances, the chairperson had no difficulty in accomplishing the interruption and did not need the advantage of priority preemption. In one session, we inadvertently gave the chairperson the lowest rather than the highest priority and she did have difficulty in carrying out her role.

As discussed above in Section 2.6, SI protocols are less satisfactory in handling collisions than SB protocols. To the extent that collisions occur during a particular test, we would expect that an SI system would be given lower ratings than a comparable SB system. We would expect that with an SI protocol which allowed all participants to be interrupters, subjects would experience more problems with collisions than would be the case with a protocol which allowed only the chairperson to be an interrupter because there would be fewer collisions in the latter case. The results of the Phase II experiments confirm this expectation and show slightly better ratings for SI systems with the chairperson as the only interrupter.

Results from SCDC experiments to be described in the next section suggest that an SI protocol which inhibited use of the interrupter path for a time long enough to allow the effects of a collision to subside would perform better than the SI protocols tested in the Phase II experiments. Such a system should handle collisions almost as well as an SB system but would allow an interrupter to be heard once a speaker had become established. Further experiments would be needed to determine the proper value for the inhibition time which unfortunately would require adjustment for different communication delays since collision effects increase in duration with those delays. We expect that SI with such modification would be marginally superior to SB without priority preemption because of the greater interruptibility of the SI protocol. However, we do not expect

that modified SI would be superior to SB with priority preemption because of the more effective interruptibility provided by preemption. Preemption is an absolute interruption heard by all the participants, but appearing as interrupter in an SI system merely means that the speaker is hearing the would-be interrupter. The speaker may choose to continue talking and ignore the attempted interruption.

2.10 Formal Experiments: Distributed-Control Techniques

SCDC conferencing techniques have been evaluated in two experimental phases. In Phase III, the performance of the four procedures for handling channel collisions was examined in detail along with the desirability of beep signals for indicating channel collisions to the participants. An SB protocol was used with a 300-msec preamble time to force all collisions to be major. In Phase IV, we used collision handling and signalling procedures indicated by the results of Phase III and examined the effects of SB, BI, and SI protocols with various preamble times. To focus attention on collisions, a new scenario called "Word Match" was developed which induced collisions at a controlled rate. In addition to subjective judgments, the scenario gave performance measurements in terms of number of word matches achieved (or tried) per unit time. PCM encoding was used in these experiments to allow comparison with an analog-bridge system having comparable overall delay.

The results of Phase III show a strong preference for signals indicating channel collision both for colliders and for listeners. In the absence of signals, collision handling procedures which favored the first speaker in a collision (procedures FS-1 and FS-2 described in Sec.2.7) were strongly preferred over the free-for-all and random suppression techniques. With these latter systems, we observed periods of tens of seconds in which nothing was heard over the channel in spite of intensive efforts to communicate by the participants which led to strong feelings of frustration. We feel that a communication system with such a property should be considered unacceptable for military use. When collision signals were used, the difference in preference judgments were much less. All procedures could be considered acceptable, but the favored-speaker techniques were still superior. There was no significant difference between the ratings of the top-ranking favored-speaker (FS-1 and FS-2) systems with signals to colliders as well as listeners. We arbitrarily chose FS-2 for use in the Phase IV experiments.

In Phase IV, SB and BI protocols were compared with three different values for preamble time: short (24 msec), long (300 msec), and extra long (1.07 sec). With short preamble times most collisions are of the minor variety which result in a short burst of noise, but otherwise there is no loss of speech. The short-preamble systems were judged to be about equal to the long-preamble systems. The extra long preamble times were judged to be significantly less satisfactory, presumably because quite a lot of speech is lost when a collision occurs with such systems.

There was no clear preference in the subjective ratings between SB and BI protocols for short and long preambles, but one error was made with the BI short system. The second channel offered by the BI protocol is not used to great advantage. Channel collisions occur with almost the same frequency since most of them follow a period of silence during which both channels are free, and the controllers in starting up all try to use the primary channel. Some reduction in channel collisions could be expected if the controllers randomly picked between the channels in situations in which both were free, but this technique was not explored because an assumed requirement of a BI protocol system would be to be capable of operating with some participants

who had only enough equipment to use the primary channel. In this regard, the experiments indicate that a mixture of SB and BI participants would not be likely to experience satisfactory conferencing. Analysis of the tapes made during the BI experiments show that between one-quarter and one-third of the utterances are carried in entirety by the secondary channel. The transmissions on the secondary channel occur because the participants overlap their utterances to some extent when they discover that this technique works for a BI system. During exchanges between two participants, we often observe a 50-50 use of the two channels. This use of overlap allows a faster interchange between the participants, but a participant who could listen only to the primary channel would miss half of the interchange.

The Phase IV experiments involved two versions of the SI protocol. SI with the SCDC communication technique is somewhat different from SI in centrally controlled conferences. Because each controller has only one receiver, the controller for a participant trying to become the channel speaker must keep its receiver set for the pseudonoise (PN) code used for the speaker channel until crypto synch has been achieved. It can then set its receiver to the PN code for the interrupter channel. When the speaker finishes talking, his controller must again switch its receiver back to the speaker's PN code in order to be able to hear the next speaker. This switching process requires some time to carry out. In the absence of detailed information about this time, we chose two values: fast (50 msec) and slow (300 msec).

Both SI systems scored well in the experiments. In fact, the SI system with slow switching received the best rating of any of the SCDC systems. However, we are convinced that the good ratings are not due to the presence of the interrupter channel. In the test with the better (slow) system, there were no complete phrases carried by the interrupter channel. The only signals on that channel were two short fragments of speech and one burst of noise caused by a collision on the interrupter channel. In the test with the fast SI system, there were one complete utterance, several fragments, and two noises. In effect, the slow SI system functions very much like an SB system with respect to collisions and does not receive the lower ratings observed with centrally controlled SI systems. However, it retains much of the interruption potential of centrally controlled SI systems. There is some reduction of effectiveness for interruption with SCDC SI because the speaker's controller stops listening to the interrupter channel as soon as he or she stops speaking. If the interrupter continues, the latter part of his utterance will appear on the speaker channel when it becomes free if there is no contention for that channel by other would-be speakers. If the interrupter's intent is to communicate with the previous speaker, a substantial part of his utterance will be lost in the switching process. If he merely wishes to stop the previous speaker and communicate with the conference as a whole, he may succeed if the speaker stops on hearing him. The ratings given the SI systems in the Phase IV tests did not reflect interruptibility because the scenario did not produce occasions when interruptions were needed.

A number of experiments were run with a delayed analog bridge system during the course of the SCDC experiments. In Phase III, the analog bridge was run at the beginning of each session to help the subjects get warmed up with respect to the word-match task and to provide them with a reference point for their ratings. In Phase IV, it was run at other times during the session to avoid any bias associated with first impressions. The ratings of the delayed analog bridge in Phase III showed it to be about equal to the better SCDC systems (those using the favored-speaker collision-handling algorithm). In Phase IV, the ratings place it near the mean of the systems tested. That some of the systems were judged to be superior to the analog bridge

is not surprising since summation is not a good technique for handling collisions. With a scenario such as Word Match, confusion can result from a collision because participants will make different interpretations from the mixture of voices. In the Phase IV experiments, the subjects made a total of four incorrect word matches (see Sec. 7.3.3). One of these errors occurred with a delayed analog bridge system. We would expect that centrally controlled systems would look even better relative to the delayed analog bridge with respect to this scenario since they do a better job of handling collisions.

To compare SCDC techniques with centrally controlled techniques, a pair of SCDC systems were included in the Phase II experiments using the consensus scenario. These systems used SB and BI protocols with short preambles and the favored-speaker (FS-1) procedure for handling collisions, but they did not have collision signals to the participants. They were rated as somewhat inferior to all centrally controlled systems except those with tandem encoding. The addition of collision signals should improve their ratings somewhat but not enough to match the better centrally controlled systems.

2.11 Conclusions and Recommendations

Our conclusions from research on voice conferencing may be summarized as follows:

(1) The requirement to accommodate narrowband encoding necessitates the use of signal selection techniques in future military conferencing systems.

(2) While signal summation (the analog bridge) is superior to all the signal selection techniques investigated, the choice of any of the better selection techniques would not result in a significant loss of conferencing capability.

(3) Speech quality (voice-encoding technique) has a larger effect on subjective judgments of system acceptability than do conferencing protocol (SB, SI, etc.) or control technique (VC, PTT, CSS).

(4) Similarly, implementation details such as the operation of speech activity detectors and the procedures for handling collisions in a SCDC conferencing system have a much greater effect than choice of conferencing protocol even though protocol questions involve much larger conceptual issues and cost considerations such as the use of additional communication channels. This sensitivity to detail means that our results cannot be considered as definitive because there is always the possibility that some other choice of implementation detail might yield a more satisfactory system. It also suggests that any procurement procedure for future systems should include simulation at a sufficient level of detail to check the implementation-dependent factors.

(5) The overall best choice of system configuration depends upon the weights given to the various factors that affect conferencing performance. These weights depend upon user requirements about which we have incomplete information. Our human-factors experiments have covered a range of

conferencing situations, and we have endeavored to design scenarios which stress conferencing capabilities and emphasize weaknesses inherent in particular techniques. No one scenario both represents what might be a "typical" military conference and serves as a good test for conferencing capability because the "typical" conference does not stress systems sufficiently to expose weaknesses. We believe that a good military conferencing system should perform as well as possible under stress (contention for the conference floor) even though military etiquette may cause most conferences to proceed with little challenge to the system.

(6) The two most important aspects of conferencing over which a system designer has some control are the ability of the system to handle collisions and the extent to which a speaker may be arbitrarily interrupted by another participant. We feel that collision handling is the more important of the two because collisions can be expected to occur far more frequently than situations in which interruption of the speaker is desirable. Collision handling can be readily assessed with test scenarios which seem relatively natural to the participants. Interruptibility is less readily assessed in an experimental situation because it is difficult to create a scenario which seems natural and also produces a reasonable number of interrupt attempts per unit of experiment time. However, the interruptibility of systems can be compared without recourse to experimentation by analyzing their control algorithms.

On the basis of our research on voice conferencing, we make the following recommendations for future secure voice conferencing systems:

(1) Voice control (VC) should be used with push-to-talk switches gating the voice signals to allow operation in noisy environments. The voice control algorithm should use a hangover time of the order of 0.4 sec to avoid rapid switching between speakers, an unsatisfactory mode of operation with narrowband encoding. If requirements for "black" controllers so indicate, the push-to-talk switches can be used to control the conference directly with little loss of conferencing performance.

(2) Centrally controlled conferencing systems should use a simplex broadcast (SB) protocol with priority preemption. The SB protocol does the best job of handling collisions, and priority preemption gives a strong interrupt capability as well as a means of recovering some conference control if a speech activity detector should fail. Priority preemption can provide a natural fit to the military rank and/or role structure in a conference. If user needs so indicate, a button could be provided to momentarily raise the priority of a participant, thereby allowing for urgent interrupts which are in conflict with the normal priority structure.

(3) Our recommendation for shared-channel distributed-control (SCDC) conferencing systems is less straightforward and depends on factors on which we do not yet have sufficient information. On the basis of present knowledge, the best choice is the speaker/interrupter (SI) protocol with slow switching that inhibits access to the interrupter channel until collisions on the speaker channel have been resolved. Collision resolution on the speaker channel should use the favored-speaker procedure described above in Section 2.7. This version of the SI protocol is as effective as the simplex broadcast (SB) protocol in handling collisions and provides some interrupt capability as well as a means of recovering some conference control if a speech activity detector should fail.

Unfortunately, the SI protocol involves some rather complex behavior on the part of the SCDC controller, and we have some doubts about the workability of the technique in practice. If synchronization problems occur in the process of switching receivers from one channel to another, participants could needlessly miss parts of the conference. We do not have sufficient information on the capabilities of the communication equipment to be used in SCDC systems to make a judgment on this question. If SCDC SI protocols should prove to be technically unsound, we would recommend the use of the SB protocol with a modification to achieve an interrupt and recovery capability. The modification would consist of a button which would force transmission on the speaker channel. Use of the button while another participant was the selected speaker would cause channel collisions that would in turn cause listeners to hear a burst of noise followed shortly by loss of crypto synchronization. On detecting loss of crypto synch, the speaker's controller would stop transmitting and signal the speaker with an audible signal or warning light to tell him that an interrupt had occurred. A participant priority structure could be used to control the use of the interrupt button, or alternatively, it could be used to allow voice control of the interrupt function resulting in a capability similar to the centrally controlled SB technique with priority preemption. Of course, if a control channel were available to allow controllers to indicate a preemption condition without having to force a channel collision, the preemption procedure could be carried out more gracefully. Any of these options for realizing a preemption capability would greatly enhance the interruptibility and recovery potential of a SCDC SB system.

The use of the broadcast interrupter (BI) SCDC protocol is not recommended. It is superior to SB with respect to allowing overlapped speech and reinforcements, but performs a little less well in handling collisions. It requires a duplicate set of costly transmit-receive and crypto equipment and an extra speech decoder. It can be argued that most of this equipment would be on hand anyway for redundancy to increase reliability. However, the later usage necessitates operation with only one

channel in the fall-back mode, and our observations indicate that conferencing would probably not be satisfactory for participants using an SB protocol in a BI conference.

(4) In high-level command and control conferences in which some participants have wideband PCM encoding for quality reasons and some participants have narrowband encoding because of communication capacity limitations, we would recommend against the use of an analog-bridge conference controller for the PCM users connected to a signal selection controller for the narrowband users. Separate interconnected controllers pose no problem, but they should both be signal selection controllers. The use of an analog bridge for the PCM users would result in difficulty for the narrowband users during any periods in which more than one PCM user was making sounds. Our experimental results do not indicate sufficient advantage for the analog bridge over the best signal selection technique to warrant its use in this situation.

3.0 OVERVIEW OF METHODS AND PROCEDURES

In this section, we present general discussions of the methods and procedures used in the experiments and of the schedule followed. These discussions are primarily intended to give the reader an overview of the conduct of the research and are not complete with respect to details of specific experimental sessions. Such details are discussed in appropriate subsections preceding the presentation of results in Sections 5 through 7.

3.1 Experimental Tasks

A number of guidelines for the design of problem-solving tasks to be employed in this research were derived from the results of searches of literature concerned with group problem solving and human information processing and the insights gained during early teleconferencing sessions. As summarized in the Interim Report on Phase I of the project,[3] these criteria were as follows:

(1) A given problem scenario should be usable over the entire range of conference sizes to be evaluated, and its difficulty should be independent of size. Furthermore, scenarios should be constructed in such a way that they can be reused with a given set of conference participants.

(2) Problems should be intrinsically interesting to subjects, and the testing situation should promote highly motivated performance.

(3) Scenarios employed should permit a variety of objective performance measures, including gross measures such as solution time and solution quality, and fine measures of communication and systems effectiveness and dynamics, such as number of messages per speaker per unit time, average queue length, and duration of pauses between messages.

(4) Scenarios should be easily learned by subjects who might differ in vocational specialty, level of formal education, intelligence, etc.

(5) Problems should be constructed in such manner that the verbal interactions required to solve them place reasonably severe demands on the capacities of each of the systems of interest. This consideration grew from observations made prior to the study that suggested that, because of the generally high quality of speech transmission in the telephone networks to be evaluated, subtle differences among systems might go undetected unless "worst case" test conditions could be devised.

(6) At least one of the scenarios should provide a context suitable for the study of military conferences where formal procedural elements such as speaker priority, chairperson control, polling, etc., would be expected to be present.

3.2 *Summary of Scenarios Developed for Teleconferencing*

With the above items as guides, a set of seven problem scenarios was developed. Descriptions of each of these, together with copies of materials used during administration, are presented in detail in Appendix B and summarized below.

3.2.1 Tasks Involving Structured Dialog

Four of the scenarios, referred to as "Number-Go-'Round," "Word-Go-'Round," "Path," and "Word Match," offer relatively straightforward assessment of the speed and ease with which information can be passed around a conference. The paradigm employed is one in which the content of the current speaker's message uniquely cues a message in the possession of one or more of the listeners. When the speaker completes his input, the listener (or listeners) cued by the input then becomes speaker and disseminates his message, cueing one (or more) other parties, etc. Three of the tasks in this set are the same except for the content of information transmitted – in one case, sequences of digits; in the second, sequences of words; and in the third, descriptions of the orientation of line segments superimposed on the cells of matrices. In the remaining scenario in the set, participants are required to exchange the contents of word lists in an effort to identify common items. In all four scenarios, participants are expected to employ fixed declarative sentence formats during their exchanges and to minimize non-solution-oriented commentaries so that task time measures can be readily analyzed.

3.2.2 Tests Involving Unstructured Dialog

Two scenarios in the group provide contexts in which participants are free to exchange task information and to produce solutions to assigned problems with few constraints on the form and content of their communication. One of these scenarios is an assignment/scheduling task in which each conference member is provided with information concerning the "home" location, "work" location, and "desired arrival time" of one or more fictitious commuters. He is also given a map showing the locations of towns identified with the problem and a listing of possible car pools that might be formed between his commuters and those assigned to other members of the conference. An experimental session begins with participants exchanging and writing information about location and times and proceeds to an interactive problem-solving phase in which conferees attempt to generate an optimal pooling and routing of commuters.

In the second scenario, the conference is presented with a brief statement of a practical problem and three or four alternative solutions. Participants then discuss the problem, the alternatives provided, and others suggested during the discussion, and attempt to reach agreement concerning the best solution.

These scenarios tend to yield a high frequency of collisions (interruptions) between speakers and provide participants with ample opportunities to interact with a given communication system both as speakers and as listeners.

3.2.3 Military Conference Scenario

The final scenario to be mentioned here combines elements from several of the scenarios identified above within a quasi-military context. Three generic military elements are simulated, a command element, a route-planning element, and a staff-support element. Participants, using various map aids, are required to exchange information concerning the condition of roads within specified geographic sectors and to construct plans that will enable troops and material to be transported in accord with tactical objectives identified during an initial command briefing. The scenario is designed to incorporate "intelligence reports" that alter the status of particular roads as viable routes and that, when delivered in the midst of a planning session, add a dynamic element to the planning.

The scenario has the advantages of being a useful tool for the study of communications within large conferences and of providing a variety of roles to be assumed by participants. Its primary disadvantage for purposes of the current evaluation is that it is relatively inefficient as a data-acquisition tool.

3.3 System and Conference Performance Measures

As indicated above in our discussion of guidelines, the possibility that differences among the teleconferencing systems might prove to be very subtle made necessary the definition of a broad set of performance measures that could provide comprehensive assessments of the ease with which the systems could be used. Three categories of measures were considered. The first category consisted of measures of total task performance, including time required to complete an assigned task or subtask, quality of task solution, and total number of alternative solutions proposed. The second contained measures of communication dynamics including number of times each participant spoke, total time each participant spoke, number of times each conferee interrupted, and was interrupted by, another speaker, average length of time each participant spent on a queue waiting for an opportunity to speak, average frequency and duration of collisions between speakers, etc. The final category contained measures of attitudes and opinions of the participants regarding the relative ease or difficulty of using the various teleconferencing systems.

3.4 Acquisition and Training of Subjects

The evolutionary development of conference capabilities and the exploratory nature of the work required a stable, experienced group of subjects who would be continually available over the course of the evaluation program. These requirements were met satisfactorily by selecting 28 persons from among approximately 32 Lincoln Laboratory volunteers in such a way as to secure the best obtainable ratios of females to males and professional to clerical staff.

For training purposes, the population of subjects was divided into subgroups, the sizes of which depended on the precise requirements of systems and scenarios to be exercised at a given time. Early in the program, each subgroup was given 5 hours of training, which included a verbal description of the system(s) to be tested and of the scenario(s) to be used, and an intensive series of practice sessions with the system(s) and scenario(s) then available. As the subjects became familiar with the scenarios, training sessions were reduced in length. By the end of the second year of work, "training" became unnecessary, and pre-session briefings were limited to verbal descriptions of the system(s) to be tested, followed by brief opportunities for practice.

Our primary goals throughout the training period were (1) to assure that subjects were thoroughly acquainted with scenario/task requirements and (2) to afford subjects ample opportunity to develop strategies for solution of the various problem types.

3.5 Design and Administration of System Evaluations

The gradual evolution of the test bed, the limited availability of experimental subjects, and the severe demands imposed by the project schedule, made necessary the adoption of a very pragmatic point of view toward system evaluation. In most instances, this point of view required that study of a given teleconferencing capability be ended, and study of another begun, as soon as a reasonable judgment concerning the efficacy of that system <u>vis-à-vis</u> earlier systems

could be made. We hoped, in most cases, to be able to make a "reasonable judgment" of acceptability on the basis of a single experimental run of the system in question.

Our approach to achieving this goal varied somewhat over the course of the evaluation. In general, however, the following statistical and procedural conventions were observed:

(1) The integrity of given subject groups was maintained whenever possible during evaluation of systems that appeared to place similar demands on conference participation. This facilitated analyses and helped to ensure that all of the information potentially available in the data could be used.

(2) Experiments containing one or more variables that could be considered, a priori, to be capable of producing strong anchor effects in the distributions of performance and preference data were conducted as a group. This was done in an effort to minimize such effects.

(3) All performance and preference data were subjected to nonparametric analyses and all conclusions concerning the statistical significance of experimental conditions were based on results obtained with nonparametric models. This choice appeared consistent with the goal of identifying the most salient effects and, in the case of subjective ratings, was mandated by the distinctly non-normal character of the data.

3.6 Schedule of Experimental Conditions

The schedule followed during evaluation was conditioned on the gradual evolution of capabilities within the test bed. During the first year, this "accumulation" provided opportunity to make some preliminary comparisons among voice-controlled speaker/interrupter, simplex broadcast, and analog-bridge systems. Further, it provided an opportunity to examine effects of conference size on performance and preference. Table 3-1 presents a summary of the (Phase I) conditions evaluated during the year.

Early in the second year, it became possible to evaluate more complex capabilities such as control-signal switching, tandeming, and certain SCDC conditions of interest. In addition, it was feasible to investigate the utilities of procedural aids such as speaker priority and chairperson control. Finally, near the end of the year, we reached the point where a wide range of possible collision-handling strategies and protocols for single- and for multiple-satellite configurations could be evaluated.

Complete summaries of the conditions evaluated during Phases II, III, and IV of the second year are presented in Tables 3-2, 3-3, and 3-4, respectively.

TABLE 3-1

CONDITIONS EVALUATED DURING PHASE I

System	Conference Size	Transmission Delay (sec)
Analog Bridge (AB)	4, 8, 12	–
Voice Control Speaker/ Interrupter (VC/SI)	8, 12	–
Analog Bridge (AB)	8	0.5
Voice Control Speaker/ Interrupter (VC/SI)	8	0.5
Voice Control Simplex Broadcast (VC/SB)	8	0.5
*CVSD Majority Voting Bridge (CVSDB)	8	–
*CVSD Simplex Broadcast (CVSD/SB)	8	–

*16 kbps. See Appendix E for description of Majority Voting Bridge.

TABLE 3-2

CONDITIONS EVALUATED DURING PHASE II

System	Delay?		Tandem?		Priority?		Chair Aids?	
	No	Yes	No	Yes	No	Yes	No	Yes
CSS/SB	X		X		X			X
CSS/SB	X		X		X		X	
CSS/SB		X	X		X			X
CSS/SB		X	X		X		X	
VC/SB	X		X		X			
VC/SI	X		X		X (SI to chair only)			
VC/SI	X		X		X (SI to all)			
VC/SB		X	X		X			
VC/SI		X	X		X (SI to chair only)			
VC/SI	X		X		X (SI to chair only)			
VC/SB	X			X	X			
VC/SB	X		X			X		
VC/SB		X	X			X		
VC/SB	X			X		X		
PTT/SB	X		X		X			
PTT/SB		X	X		X			
PTT/SB	X		X			X		
PTT/SB		X	X			X		
PTT/SI	X		X		X (SI to chair only)			
PTT/SI	X		X		X (SI to all)			
*SCDC/BI	X		X		X			
*SCDC/SI	X		X		X			
VC/SB(20)	X		X		X			
VC/SI(20)	X		X		X			

*SCDC conferences utilized distributed-control procedures. All other conferences utilized central-control procedures. (See Sec.2 for explanation of differences.)

TABLE 3-3

CONDITIONS EVALUATED DURING PHASE III
(Fixed parameters: system = PTT, protocol = SB, preamble = 300 msec, encoding = PCM)

Speaker Selection Strategy	Signal to Collider Only	Signal to Collider and Listeners	No Signal
First speaker: Version 1	X	X	X
First speaker: Version 2		X	X
Free-for-All	X	X	X
Random Suppression	X		X

TABLE 3-4

CONDITIONS EVALUATED DURING PHASE IV
(Fixed parameters: system = PTT, encoding = PCM, collision strategy = first speaker: Version 2)

Protocol	Preamble Time (msec)
Simplex Broadcast (SB)	24, 300, 1067
Broadcast/Interrupter (BI)	24, 300, 1067
Speaker/Interrupter with Fast (50 msec) Switching (SIF)	24
Speaker/Interrupter with Slow (300 msec) Switching (SIS)	24

4.0 PHASE I

NOTE: In order to simplify the discussion of results of the various phases in the series of experiments, we shall divide our presentation of each phase into subsections, as follows:

.1 Statement of Purpose
.2 Summary of Procedure
.3 Presentation of Results

Further, in an effort to avoid duplication of descriptive information, particularly with respect to procedures that are common across conditions, we shall attempt to limit discussion to unique characteristics of a given phase.

4.1 Statement of Purpose

The first phase of the evaluation study served a variety of purposes. Prior to its inception, some informal experience had been gained by Lincoln and BBN staff with the prototype voice control teleconferencing systems then available in the test bed, and a number of brief pilot studies had been run with project personnel in an effort to improve characteristics of the scenarios. The limit of what could reasonably be expected in the way of knowledge concerning engineering of the systems and design and administration of relevant experiments had been approached during this time, and it was now appropriate to begin more formal study. The first goals to be identified with Phase I, then, had to do with accumulation of information on a number of dimensions: (1) reliability of test-bed systems; (2) responsiveness of scenarios and measurement procedures; (3) evaluation of subject recruitment, training, and briefing procedures; and (4) assessment of measurement techniques.

In addition to these goals, which, though important, were certainly not unique to this particular project, was a series of goals associated with test and evaluation of specific teleconferencing variables and systems identified in the project work statement. In terms of the actual experimental comparisons that were finally made, these variables were as follows:

(1) Evaluation of the effects of conference size on conference performance. The goal was to compare conferences of 4, 8, and 12 participants with respect to speed and quality of performance, amount of speech generated, and number and rate of collisions and interruptions. Since all assessments relating to conference size were made with the analog bridge system, comparisons made here also served the purpose of providing baseline data against which to compare results obtained with other voice protocols.

(2) Comparative evaluation of an analog bridge and a voice-controlled/speaker-interrupter system in medium (8-participant) and large (12-participant) conferences.

(3) Comparative evaluation of an analog bridge, a voice-controlled simplex broadcast, and voice-controlled/speaker-interrupter system in environments that included delays similar to those experienced in satellite communications (0.5 sec).

(4) Comparison of a CVSD simplex broadcast system with a CVSD bridge system. This comparison provided the only opportunity during the phase to experiment with systems with reduced intelligibility.

It was recognized from the outset that, because of the complexities of problems involved in effectively integrating new hardware, software, and procedures, and in maintaining a group of trained and dedicated conference participants, the above goals could only be approximated in the time available. Nonetheless, it was expected that sufficient information could be gathered to enable at least preliminary conclusions to be drawn regarding the relative efficacies of Phase I teleconferencing arrangements.

4.2 Summary of Procedure

Five separate experimental comparisons were made during Phase I. A schedule of these comparisons appears in Table 4-1.

TABLE 4-1

SCHEDULE OF PHASE I COMPARISONS

Comparison	Conference Size(s)	System(s)	Special Conditions
1	4 vs 8	Analog Bridge (AB)	–
2	8	AB vs Voice-Controlled Speaker/Interrupter (VC/SI)	–
3	12	AB vs VC/SI	–
4	8	AB vs VC/SI vs Voice-Controlled Simplex Broadcast (VC/SB)	0.5 sec (satellite) delay
5	8	CVSD Majority Voting Bridge vs CVSD Simplex Broadcast	–

Each experimental session was divided into two $\frac{1}{2}$-hour periods. At the beginning of each period, the experimenter conducted a short briefing which included a description of the teleconferencing system to be used during that period, the locations of telephones, and telephone numbers to be used by conferees. He then answered questions, distributed materials required for solution of the conference scenario, and selected one person to serve as a "starter" for the session. Following this, conferees were released to locate their telephones, to initiate the dial-up procedure, and to practice with the system.

When the starter had verified that all persons had entered the conference and were able to communicate successfully with each other, the experimenter gave a signal to begin. The starter informed the rest of the conferees that the signal had been given and the session was initiated. At this point, the experimenter started a tape recorder and began monitoring the proceedings of the conference with the aid of headphones.

When, in the judgment of the experimenter, conferees had reached consensus that the best solution to the experimental problem had been found, or a period of 18 min. had elapsed since the "start" signal, the starter was instructed to inform conferees that 2 min. remained before termination of the session. When this latter period of time had elapsed, conferees were advised that the conference was over and that they should return to the main conference room.

When all had returned, a debriefing session was held. During this session, subjects were told whether or not they had achieved the optimal solution to the problem and, if not, what the optimal solution was. In addition, comments were solicited on the voice quality of the communication lines, special difficulties associated with interrupting other speakers or being interrupted by them, and any other factors pertinent to use of the teleconferencing system. When the debriefing session was complete, orientation for the next session commenced or the subjects were dismissed, depending on which half-hour period had just been completed.

Beyond these elements of procedure, which were constant across the five experiments, some critical differences existed among the various comparisons with respect to administration of test scenarios. These differences are dealt with in separate subsections below.

4.2.1 Comparison I: Four-Versus Eight-Person Analog Bridge Conferences

The first of the studies conducted had two primary goals: (1) evaluation of the effects of conference size on performance and (2) verification of the assumption that the car-pool scenario met our criterion that problem difficulty should be independent of problem size.

To satisfy these goals, two equivalent versions (transforms) of each of four eight-commuter car-pool problems were generated. Four unique groups of eight conferees were chosen randomly from the pool of trained subjects, and each was paired with one of the four problems. A given combination of conferees then solved one version of its problem in a single full conference containing eight persons, and the second version in two independent conferences containing four persons each. To control against the possibility of sequence effects, persons in two of the groups participated first in the larger conference and then in the smaller ones, while those in the remaining groups participated first in the smaller conferences.

4.2.2 Comparison II: Eight-Person Analog Bridge and VC Speaker/Interrupter Conferences

For purposes of comparing performance in the Analog Bridge system with that in the VC Speaker/Interrupter System, two unique groups of eight conferees were drawn from the subject pool and each was given an eight-commuter problem similar to that utilized in Experiment I. Although the schedule permitted only a single replication of the comparison, efforts were made, as before, to control sequence effects by counterbalancing the order in which the groups were exposed to the conferencing conditions.

4.2.3 Comparison III: Twelve-Person Analog Bridge VC Speaker/Interrupter Systems

The growth of the test-bed facility to the point where conferences containing 12 persons could be supported provided an opportunity to evaluate the VC Speaker/Interrupter System in the context of a moderately large conference. As in earlier studies, the Analog Bridge System served as a control condition against which to compare performance.

Four conference groups involving as many unique combinations of participants as possible within the constraints of subject pool size and work schedule were formed for this experiment.

Each group was assigned one of four 12-commuter problems. As in Experiment I, two versions of each of these problems were prepared and exposure of the conference groups to system conditions was counterbalanced.

4.2.4 Comparison IV: Eight-Person Delayed Analog Bridge, VC Speaker/Interrupter, and Simplex Broadcast Conferences

The effects of adding a 0.5-sec delay between the origination of speech and its receipt by listeners (similar to the delay that would be experienced by conferees communicating via satellite) were studied in the experiment. Each of the three types of systems was paired with each of the remaining systems, yielding a set of three unreplicated comparisons. Three unique groups of 8 conferees, drawn from the pool of 14 subjects then available, solved transforms of 8-commuter problems utilized earlier in the series.

4.2.5 Comparison V: Eight-Person CVSD Majority Voting Bridge and Simplex Broadcast Conferences

The final experiment in this series was concerned with a comparison of performance in a CVSD system (App. E) that permitted listeners to hear all speakers simultaneously engaged in speaking, with performance in a CVSD system that permitted only one speaker to be heard at a given time. Two equivalent eight-commuter problems utilized in Experiment I were selected for this comparison, and a single group of eight conferees was selected from the subject pool. The first problem was solved on the CVSD Bridge; the remaining problem was then solved on the Simplex Broadcast System.

4.3 Methods of Analysis

The car-pool task designed for use in this research provided relatively direct means for assessing quantitative and qualitative aspects of total conference output. The following specific measures were selected for use with the car-pool problem: (1) best score actually achieved during an experimental session, to be compared with the theoretically optimal score; (2) time required to achieve the best score; and (3) time required to achieve the first complete allocation.

In addition to these gross measures of conference performance, a number of measures of the fine structure of a conference were defined. Compilation of data supporting these measures was accomplished by careful auditing of each of the tape recordings made during the evaluation, in accord with conventions identified below.

4.3.1 Total Speech Time

An estimate of total speech time was made by summing the durations of all speech energy segments detected by the listener over the course of a session. A segment was considered to have begun, and a time clock was started, when energy was first detected; it was considered to have ended, and the time clock was stopped, when no further energy could be detected.

Because of the likelihood of timing errors during very rapid exchanges between speakers, no effort was made during the timing of speech segments to distinguish voices. Thus, a given speech segment in this analysis might consist of energy supplied by a single speaker, or of the energies supplied by two or more speakers whose voices were heard in very rapid succession.

Note that, because this procedure does not distinguish situations in which only one conferee is speaking from those in which several conferees are speaking simultaneously, it leads to a

measure of total speech time that may occasionally underestimate the amount of speech that can actually be heard in conferences employing a bridge system.

4.3.2 Interrupting Speech Not Heard by Conference

Estimates of the total duration of speech energy produced by conferees not selected to be speakers in Simplex Broadcast and Speaker/Interrupter system conferences were made by accumulating speech energy segments detected on that track of the recording associated with activity in the interrupter channel. The duration of a segment was assessed without differentiating among "interrupting" voices. A ratio of total duration of interrupting speech to total duration of speech, derived as above, was then computed for each Speaker/Interrupter session. The set of ratios is presented in the tables to follow.

This procedure and the one to be discussed in the next section would be expected to yield occasional underestimates of the actual frequencies and durations of attempted interruptions when several conferees speak simultaneously.

4.3.3 Average Time Between Interruptions

An estimate of the average time between interruptions was made by dividing the total speech time associated with a given conference by the total number of interruptions that could be detected during that time. An interruption was defined as any instance in which two or more conferees appeared to be speaking simultaneously. As earlier, no effort was made to differentiate among speakers who had produced an interrupt event.

The procedure used to estimate average time between interruptions does not distinguish random "collisions," in which two conferees begin to speak simultaneously after a period of silence, from intentional interruptions of one conferee by another. This lack of distinction is considered to be of no great concern in the current context. It is important to note, however, that it would be impossible, on the basis of such an estimate alone, to decide whether a change in interruption rate observed as the result of manipulation of a given conferencing variable (e.g., conference size) was due simply to a change in the frequency of unavoidable "collisions," to a change in conferee willingness to interrupt, or to both.

4.3.4 Information Acquired via Questionnaire

After Experiment III, and again after Experiment IV, participants were required to fill out short questionnaires. The most important item on these questionnaires required an estimate of the relative ease or difficulty of using the various teleconferencing systems studied up to this time. To complete this item, 12 participants marked the position of each system on a scale that ran from "difficult" to "easy" in accord with their perception of the use of the system in question. The resulting scales provided reasonably accurate indications of the rank order and relative magnitudes of ease of use.

On the later of the questionnaires, participants were asked to make several additional ratings on dimensions related to perceived ease of interruption of a speaker, ability to recognize speakers, and estimated ability to be heard and recognized by listeners. Ratings on these dimensions were used during interpretation of responses made to the overall ease-of-use item discussed immediately above.

TABLE 4-2

SUMMARY OF SCORES AND SOLUTION TIMES ACHIEVED BY FOUR- AND EIGHT-PERSON CONFERENCES USING THE ANALOG BRIDGE SYSTEM

Problem	Measure	Four Participants		Eight Participants
		Group 1	Group 2	Group 1 + Group 2
P_1	Score	111	109*	110
	Time to first	1.05	1.30	3.08
	Time to best	1.05	15.90	19.78
P_2	Score	123*	123*	123*
	Time to first	1.56	2.02	1.9
	Time to best	1.56	10.92	6.88
P_3	Score	124*	124*	124*
	Time to first	1.52	1.24	2.2
	Time to best	3.72	1.24	7.0
P_4	Score	88*	88*	88*
	Time to first	2.58	6.77	1.75
	Time to best	2.58	6.77	2.59

*Denotes actual score equal to theoretically optimal score based on linear program.

TABLE 4-3

PERCENTAGES OF INFORMATION-DISSEMINATION AND PROBLEM-SOLUTION PHASES ACTUALLY SPENT IN COMMUNICATION BY FOUR- AND EIGHT-PERSON CONFERENCES USING THE ANALOG BRIDGE SYSTEM

Problem	Phase	Four Participants		Eight Participants
		Group 1	Group 2	Group 1 + Group 2
P_1	Information Dissemination	67.5	56.0	69.1
	Problem Solution	53.9	44.0	66.8
P_2	Information Dissemination	36.4	38.3	67.9
	Problem Solution	52.4	36.1	56.6
P_3	Information Dissemination	81.5	45.3	66.7
	Problem Solution	63.0	31.1	55.5
P_4	Information Dissemination	70.3	72.6	69.7
	Problem Solution	46.4	58.2	62.4

4.4 Results

4.4.1 Experiment I: Four- Versus Eight-Person Analog Bridge

1. Total Conference Performance

Table 4-2 presents results obtained with conferences of four and of eight persons using the Analog Bridge system. The results are tabulated for three measures of total conference performance: (1) best score achieved, (2) time elapsed from end of information dissemination period to formulation of first solution, and (3) time elapsed from end of information dissemination period to formulation of the best solution achieved during the experimental session. As explained earlier, each of the problems utilized during this portion of the study was solved by eight persons working as two independent teams of four and as a single team of eight, hence the column identifiers, "Group 1," "Group 2," and "Group 1 + Group 2."

Several observations can be made with respect to the data contained in this table. First, optimal scores were achieved in all but two instances (P_1, Group 1, and Group 1 + Group 2), and, even in those instances, performance was only marginally suboptimal.

Second, in a surprisingly large percentage of cases (42 percent), the first solution achieved by a conference was the best achieved by it over the course of a session. All of these "best-first" performances were produced by conferences of four persons.

Finally, there is no systematic difference between the two conference sizes with respect to the amount of time required to produce either the first or the best solution.

2. Amount of Speaking Time

The percentages of conference times that four- and eight-person groups actually spoke during the information dissemination and problem-solving phases are presented in Table 4-3. Once again, there appears to be no relationship between conference size and output.

Within the constraints on accuracy associated with this measure of speaking time, we conclude that increasing conference size from four to eight persons does not produce a systematic change in the amount of speech generated.

3. Time Between Interruptions

Average times between interruptions during four- and eight-person conferences are presented in Table 4-4.* Note here that, with one exception (P2, Group 2 vs Group 1 + Group 2), these averages are greater for conferences of four, indicating a lower rate of interruption in these conferences.

4. Conferee Attitudes

Attitudes and opinions of participants obtained informally at the conclusion of each session in this series suggested little in the way of tangible differences between four- and eight-person conferences with respect to difficulty of problem solution. Most persons agreed, however, that both single and multiple interruptions of a given speaker were more frequent in the larger conferences. These conferees reported adoption of a sort of "self-discipline" in an effort to minimize the frequency of such "collisions." One of the characteristics of this discipline that could be clearly identified in the responses was a requirement for a longer pause on the part of a current speaker before his interruption by a listener waiting to speak.

---

* Interruptions are not to be expected, and were judged to be of very low frequency, during the information dissemination phase. As a result, they are not presented in any of the tables in this section.

TABLE 4-4

AVERAGE TIME (in sec) BETWEEN INTERRUPTIONS IN FOUR- AND EIGHT-PERSON CONFERENCES USING THE ANALOG BRIDGE SYSTEM

Problem	Four Participants		Eight Participants
	Group 1	Group 2	Group 1 + Group 2
P_1	13.82	18.74	11.66
P_2	16.58	11.19	12.52
P_3	11.25	10.22	9.40
P_4	15.55	18.75	8.62

TABLE 4-5

SUMMARY OF SCORES AND SOLUTION TIMES ACHIEVED BY EIGHT-PERSON CONFERENCES USING ANALOG BRIDGE AND VC SPEAKER/INTERRUPTER SYSTEMS

Problem	Measure	Analog Bridge	VC Speaker/Interrupter
P_5	Score	128*	128*
	Time to first	1.90	3.75
	Time to best	9.32	15.92
P_6	Score	120*	120*
	Time to first	2.25	3.54
	Time to best	2.25	3.54

*Denotes actual score equal to theoretically optimal score based on linear program.

4.4.2 Experiment II: Eight-Person Analog Bridge and VC Speaker/Interrupter Conferences

Table 4-5 presents the results obtained with conferences of eight persons using the Analog Bridge and VC Speaker/Interrupter systems. As the scores indicate, optimal solutions were found for both problems under both systems. The times required to produce the first solution in each session are slightly less under Analog Bridge conditions, but the times required to reach best solutions are randomly distributed in this small sample.

In summary, there is nothing in these data to suggest that, with respect to total conference performance, any difference exists between the two systems.

1. Amount of Speaking Time and Time Between Interruptions

Percentages of total problem solving time that conferees actually spoke, and average times between interruptions are presented in Table 4-6. As indicated earlier, speech that occurs within the VC Speaker/Interrupter System can be categorized as having one of two fates, depending upon who originates it. If it is originated by the designated speaker, it reaches the floor of the conference and can be heard by all listeners. If it is originated by an interrupter, it reaches only the designated speaker. This distinction has been maintained in the organization of the table.

TABLE 4-6

COMMUNICATION PERCENTAGES AND AVERAGE TIME BETWEEN INTERRUPTIONS IN EIGHT-PERSON CONFERENCES USING ANALOG BRIDGE AND VC SPEAKER/INTERRUPTER SYSTEMS

Problem	Measure	Analog Bridge	VC Speaker/Interrupter
P_5	Percent of Total Time Speech Occurred and Was Heard by Conference	44.6	41.1
	Percent of Total Time Interrupting Speech Occurred and Was Heard Only by Speaker	N/A	*
	Time Between Interruptions (sec)	7.5	*
P_6	Percent of Total Time Speech Occurred and Was Heard by Conference	30.8	34.44
	Percent of Total Time Interrupting Speech Occurred and Was Heard Only by Speaker	N/A	23.88
	Time Between Interruptions (sec)	17.2	2.85

*Recording failure.

TABLE 4-7

SUMMARY OF SCORES AND SOLUTION TIMES ACHIEVED BY TWELVE-PERSON CONFERENCES USING ANALOG BRIDGE AND VC SPEAKER/INTERRUPTER SYSTEMS

Problem	Measure	Analog Bridge	Voice Control
P_5	Score	174	174
	Time to first	2.45	1.66
	Time to best	2.45	9.83
P_6	Score	178	174
	Time to first	4.54	7.1
	Time to best	6.3	7.1
P_7	Score	174*	174*
	Time to first	3.4	4.4
	Time to best	5.0	4.4
P_8	Score	186*	186*
	Time to first	3.5	4.08
	Time to best	10.02	8.48

*Denotes actual score equal to theoretically optimal score based on linear program.

TABLE 4-8

COMMUNICATION PERCENTAGES AND AVERAGE TIME BETWEEN INTERRUPTIONS IN TWELVE-PERSON CONFERENCES USING THE ANALOG BRIDGE SYSTEM

Problem	Phase	Percentage of Time Spent in Communication	Time Between Interruptions (sec)
P_5	Information Dissemination	55.1	N/A
	Problem Solution	38.7	4.44
P_6	Information Dissemination	36.4	N/A
	Problem Solution	55.0	5.39
P_7	Information Dissemination	46.0	N/A
	Problem Solution	53.1	7.83
P_8	Information Dissemination	35.0	N/A
	Problem Solution	54.3	8.89

Although the data presented here are too few for purposes of establishing statistical significance, two observations may be of interest: (1) the percent of total conference speech time that interrupting speech occurs and cannot be heard by the conference at large appears to be quite substantial; (2) the rate at which interruptions occurred in the VC conference is very much higher than in the Analog Bridge conference. If these outcomes prove to be reliable in the face of replication of the experiment, they may underscore the need for procedures that could prevent the possible loss of critical information contributed during uncontrolled interruptions of designated speakers (e.g., buffering the interrupting speech until it could be introduced after a selected speaker had relinquished the floor).

3. Conferee Attitudes

Attitudes and opinions collected informally during this experiment indicated that the VC system was more difficult to use than the Analog Bridge system. Participants agreed that when they had been selected to be speakers and were addressing the conference, they found the occurrence of an interruption, which they knew could not be heard by the conference at large, to be disconcerting. The reason for this appeared to be that speakers felt they had to attend to the interrupter's speech more closely than they felt they needed to while using the Bridge. Presumably, this requirement to divide attention interfered with messages they, as speakers, were attempting to input. Some conferees also reported that they occasionally noted lapses and evidences of indecision on the part of speakers which they attributed to speakers' listening to interrupters.

The final point was made that it was much more difficult to gain the floor with the VC system. Conferees felt, in general, that they had to make more frequent and concerted efforts to secure the floor, an observation that we believe is corroborated by the high interruption rate implied in Table 4-6.

4.4.3 Experiment III: Twelve-Person Analog Bridge and VC Speaker/Interrupter Conferences

1. Total Conference Performance

Table 4-7 presents a summary of scores and solution times achieved by 12-person conferences using the Analog Bridge and the VC Speaker/Interrupter systems. The same high quality of performance noted in connection with the smaller conferences of Experiment I is found here, although, as one would expect, more time tends to be required to produce "first" and "best" solutions with the 12-commuter problems than with the 8-commuter problems.

On the basis of the performance scores and times reported in Table 4-7, we conclude that there are no significant differences between the two conference systems under the conditions studied.

2. Amount of Speech and Average Time Between Interruptions Using Analog Bridge and VC Speaker/Interrupter Systems

Estimates of the percentage of speech during information-dissemination and problem-solution phases of the car-pool problems, and estimates of the average times between interruptions during the latter phase are presented in Tables 4-8 and 4-9.

TABLE 4-9

COMMUNICATION PERCENTAGES AND AVERAGE TIME BETWEEN INTERRUPTIONS IN TWELVE-PERSON CONFERENCE USING THE VC SPEAKER/INTERRUPTER SYSTEM (Problem Solution Phase Only)

Problem	Percent of Total Time Speech Occurred and Was Heard by Conference	Percent of Total Time Interrupting Speech Occurred and Was Heard Only by Speaker	Mean Time Between Interruptions (sec)
P_5	30.0	10.0	1.47
P_6	27.6	4.7	2.16
P_7	30.8	4.5	2.74
P_8	42.0	4.0	3.30

TABLE 4-10

SUMMARY OF SCORES AND SOLUTION TIMES ACHIEVED BY EIGHT-PERSON CONFERENCES USING DELAYED ANALOG BRIDGE, DELAYED VC SIMPLEX BROADCAST, AND VC SPEAKER/INTERRUPTER SYSTEMS

Measure	Analog Bridge With Delay	VC Simplex Broadcast With Delay	VC S/I With Delay
Score	114		113
		120*	128*
	113	120*	
Time to First Solution (min.)	3.0		10.0
		2.37	1.74
	2.22	2.82	
Time to Best Solution (min.)	11.8		13.75
		2.37	8.7
	4.5	7.3	

*Denotes actual score equal to theoretically optimal score based on linear program.

A comparison of these tables indicates the following:

(1) Considerably less of the speech generated by conferees using VC reached the floor of the conference than did that of conferees using Analog Bridge.

(2) The average time between interruptions was consistently lower during solution of problems over VC.

Table 4-9 also contains estimates of the amount of interrupting speech that occurred and could have been detected only by the selected speaker (column three). Given that the quality and pace of performance was approximately equal to that observed with the Analog Bridge, it seems likely (and the comments of conferees suggest strongly) that the lost speech was not essential to the business of the conference.

3. Conferee Attitudes

Conferees consistently registered strong preferences for the Analog Bridge system during debriefing sessions. They reported the realization that they could, as listeners waiting to speak in the VC system, decide rather easily whether or not to attempt an interruption, but that process interfered with their thoughts concerning what they had to say. Most conferees reported that they intentionally added preambles to their statements (e.g., "This is Marge and I want to say ...") in an effort to ensure that early portions of their inputs, which might be lost, would not contain information critical to the proceedings. (These reports were verified during the analysis of the tapes.) Application of this strategy, though largely successful in securing the floor at little cost in information, was considered by the conferees to be a nuisance.

4.4.4 Experiment IV: Eight-Person Delayed Analog Bridge, VC Simplex Broadcast, and Speaker/Interrupter Conferences

1. Total Conference Performance

Table 4-10 presents the scores and solution times associated with the three systems studied. As indicated in Section 3.5.4, no attempt was made to control order of the presentation of car-pool problems in this experiment. However, the problems were considered to be equivalent in difficulty and would be expected to lead to similar performance scores if the teleconferencing systems were equally easy (or difficult) to use.

The lack of replication of comparisons presented here prevents the drawing of any conclusions related to total conference performance. It does seem clear, however, that the scores and times compare favorably with those presented for eight conferees in Table 4-5. We are inclined to believe that, though consistent differences may be uncovered during later replication, they are not likely to be of practical significance.

2. Amount of Speech and Average Time Between Interruptions

The percentages of total conference time that speech occurred and average times between interruptions for the three conditions are presented in Table 4-11. As above, we are unable to draw any conclusions regarding these estimates, because of lack of replications. Once again, however, it is interesting to compare the values tabulated for the delayed Analog Bridge against those obtained with the Analog Bridge of Comparison I (Table 4-4). The sizes of the differences in both amount of speech and average time between interruptions suggest a rather strong effect due to the simulated delay, and, in our judgment, deserve further study.

TABLE 4-11

COMMUNICATION PERCENTAGES AND AVERAGE TIME BETWEEN INTERRUPTIONS IN EIGHT-PERSON CONFERENCES USING DELAYED ANALOG BRIDGE, DELAYED VC SIMPLEX BROADCAST, AND VC SPEAKER/INTERRUPTER SYSTEMS

Measure	Analog Bridge With Delay	VC Simplex Broadcast With Delay	VC Speaker/Interrupter With Delay
Percent of Total Time Speech Occurred and Was Heard by Conference	31.57		28.15
		28.44	29.27
	32.00	27.75	
Percent of Total Time Interrupting Speech Occurred and Was Not Heard by Conference (or Was Heard Only by Speaker)	N/A		(6.11)
		*	(19.01)
	N/A	9.51	
Time Between Interruptions (sec)	4.98		10.29
		N/A	6.78
	9.13	N/A	

*Time not obtainable in this analysis of tapes.

TABLE 4-12

SUMMARY OF SCORES AND SOLUTION TIMES ACHIEVED BY EIGHT-PERSON CONFERENCES USING THE CVSD BRIDGE AND CVSD SIMPLEX BROADCAST SYSTEMS

Measure	CVSD Bridge	CVSD Simplex Broadcast
Score	128*	116
Time to First Solution (min.)	2.0	1.9
Time to Best Solution (min.)	2.0	6.05

*Denotes actual score equal to theoretically optimal score based on linear program.

3. Conferee Attitudes

Conferees reported that they were very aware of the delay introduced into these systems, particularly that associated with Speaker/Interrupter and Simplex Broadcast. They felt that the delay presented an initial impediment to the free flow of conversation, but that, by slowing the pace of the conference slightly, the effect could be overcome. Most agreed that the major problem experienced during the sessions was occasional inability to determine whether they were being heard. They compensated for this by repeating inputs "to be sure of getting through." Finally, the conferees reported that they found it more difficult to interrupt speakers during these conferences than during earlier (undelayed) conferences.

4.4.5 Experiment V: Eight-Person CVSD Majority Voting Bridge and CVSD Simplex Broadcast Conferences

1. Total Conference Performance

Performance scores and problem solution times associated with eight-person conferences using the CVSD Bridge and Simplex Broadcast systems are presented in Table 4-12. Although no conclusions can be drawn from this single experimental session, the performance represented here does not appear dissimilar to that identified with eight-person conferences using the Analog Bridge and VC systems.

2. Amount of Speech and Average Time Between Interruptions

Percentages of total conference time spent speaking and the average time between interruptions observed with the CVSD Bridge are presented in Table 4-13. It is interesting to note that

TABLE 4-13

COMMUNICATION PERCENTAGES AND AVERAGE TIME BETWEEN INTERRUPTIONS IN EIGHT-PERSON CONFERENCES USING THE CVSD BRIDGE AND CVSD SIMPLEX BROADCAST SYSTEMS

Measure	CVSD Bridge	CVSD Simplex Broadcast
Percent Communication	43.21	24.77
Time Between Interruptions	5.20	N/A

the value associated with percent communication in the Simplex Broadcast system is the lowest observed over the course of the project.

3. Conferee Attitudes

Conferees reported that they found both CVSD systems "unpleasant" to use, though they felt their overall performance was probably equal to that in other experiments. All agreed that the quality of speech was inferior in these systems. Transmissions were punctuated by spurious noises and occasionally words or speakers were not recognized. Of the two systems, the CVSD Bridge seemed the more difficult to use, and conferees felt a greater need to repeat their messages to be certain that they were understood while using this system.

In considering the attitudes and opinions of conferees with respect to the CVSD Bridge and Simplex Broadcast, it is important to bear in mind that these systems were the only ones studied during the series that could be characterized as having reduced intelligibility and low signal-to-noise ratio.

4.5 Summary of Phase I

In this section, we have discussed five studies concerned with the effects of number of conferees, type of teleconferencing system, and transmission delay on conference performance. Although much more research would be required before the results of the studies could be considered valid, the outcomes of certain experimental comparisons are compelling and deserve comment here.

4.5.1 Effects of Conference Size

Increasing the number of conferees from four to eight appears to have little effect on the quality and pace of problem solving in Analog Bridge conferences. The results obtained with 12-person conferences using this system, though indicating a slower conference pace, are very similar to those obtained with the smaller conferences. In the aggregate, the data suggest that conferences of 4, 8, and 12 persons cannot be distinguished from each other on the basis of gross measures of quality and productivity.

Important differences may exist, however, with respect to the rates at which attempts are made to interrupt speakers. The data suggest a progressive decrease in the average time elapsed between interruptions over the range of conference sizes studied. Further, the conferees report an awareness of an increase in the frequency of interruptions, and attempt to compensate by waiting for longer pauses in the speech of a given speaker before attempting interruption. The success of this strategy cannot be assessed in absolute terms with our current methodology, but it seems clear that the interruption rates for 12-person and 8-person conferences remain higher than those for 4-person conferences.

Finally, the comments of subjects suggest that conferencing becomes more difficult as conference size increases. This increasing difficulty may be due to the need for adoption of strategies such as that mentioned above.

4.5.2 Effects Due to Type of Conferencing System

Our studies suggest that there are no significant differences among the Analog Bridge, VC, and CVSD Bridge systems with respect to gross measures of conference output. For conferences of the types and sizes examined, all systems, including the VC Speaker/Interrupter and VC and CVSD Simplex Broadcast could be expected to provide sufficient bandwidth for the accomplishment of group problem-solving tasks.

As in the case of conference size, one must look to fine measures of conferees' interactions and to the comments of the conferees to distinguish among systems. Our best measure, average time between interruptions, suggests rather strongly that the rate of interruptions is significantly higher in conferences using VC and CVSD than in those using the Analog Bridge, and we have pointed out several possible reasons for this finding. Unfortunately, it is impossible to scale the various versions of VC and CVSD studied here with respect to interrupt rate because of the small amount of data taken and limitations in our current ability to measure frequencies of attempted interruption in the simplex broadcast versions of those systems.

The comments of conferees indicate an awareness of higher frequencies of interruptions in the VC and CVSD conferences. The need to cope with this increased frequency and with the occasional loss of transmissions that occurs in Simplex Broadcast and Speaker/Interrupter modes creates a more difficult conferencing environment than that associated with the Analog Bridge. Among those we studied, however, only the CVSD Bridge system comes close to being unacceptable to conferees.

4.5.3 Effects of Delay

Aside from a possible slight reduction in conference pace, we are unable to find any impact on total performance caused by the introduction of a 0.5-sec delay in the transmission of speech. On the basis of our study, we are inclined to believe that the existence of satellite delays of this duration will produce no effect on the general quality and productivity of a teleconference.

The effect on interaction of a delay of this magnitude is clearly perceived by the conferees. They report deliberate efforts to slow the pace of their transmissions, to repeat their messages, and to limit the frequencies of their interruptions in order to maintain conference quality. The results of the single comparison between delayed Analog Bridge and delayed VC Speaker/Interrupter performed here suggest that these efforts are reasonably successful.

4.5.4 Overall Assessments of Ease of Use

The two questionnaires completed by participants provide what is perhaps the most concise summary available concerning actual use of Phase I systems (excluding VC-Simplex Broadcast and CVSD). The distributions of relative ease of conferencing in various conditions, as estimated from the responses to the questionnaire item concerned with overall system rating are presented as a final footnote to this phase of the work. For purposes of presentation, the scale generated by each participant was normalized by computing the ratio of the distance of each scale marking from the nominal zero position ("hard") to the total length of scale utilized. Means of these normalized scales are presented in the figure.

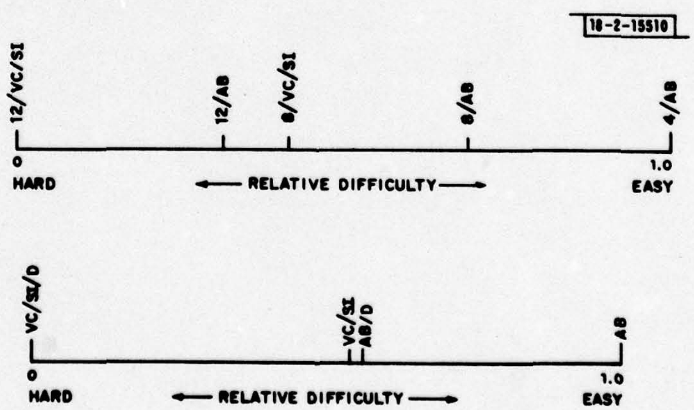


Fig. 4-1. Mean normalized estimates of relative difficulty of using Voice Control/Speaker Interrupter (VC/SI) and Analog Bridge (AB) systems with and without delay (D). Numbers preceding slashes indicate conference size.

5.0 PHASE II

A significant observation the first year was that conference participants were able to report, and to agree on, differences in the amounts of effort required to use various systems in instances where more direct measures failed to find differences in performance. This observation, reminiscent of that of Richards and Swaffield (1958; see App. A), suggested that more formal efforts should be made to acquire attitude and judgment data. It suggested further that a more significant role be assigned to these data in the comparative evaluation of systems. The view was justified on the pragmatic grounds that, if the scenario(s) successfully captured those aspects of real-world teleconferencing environments of interest to system designers, yet measures of performance failed to discriminate among system alternatives, the only basis for choice might lie in information derived from subjective ratings. For purposes of acquiring ratings, a new questionnaire was designed and used as a primary source of information throughout the remaining phases of the study. This questionnaire is described in Section 5.2.3.

The experience gained during Phase I with respect to the "Car Pool" scenario was also illuminating. As its detailed exposition in Appendix B suggests, this scenario met most criteria established for scenario design at the beginning of the project. In addition, the formal mathematical problem it posed could be solved by linear programming, yielding an optimal score for comparison against actual performance.

From a methodological point of view, however, the scenario was deficient in several important respects. Chief among these were: (1) It required a considerable investment in training time. (2) It could not be understood sufficiently well by several of the initial volunteers. (3) Perhaps, most importantly, it was inefficient as a generator of data. It seemed clear that further comparisons among systems could be (indeed, in view of the schedule, would almost have to be) conducted using simpler, shorter tasks than the one which served us in Phase I.

The "Number-Go-'Round" and "Path" tasks designed earlier but essentially unused during Phase I served as models for scenarios designed for Phase II and used throughout most of the remaining work. These scenarios, "Word-Go-'Round" and "Word Match," are discussed in detail in Appendix B.

A third scenario, "Consensus," was also designed at this time. In this scenario, a brief description of a problem or dilemma is read to the conference which then attempts to reach consensus on a solution or course of action. The scenario has a number of distinct advantages over others used during the series. It requires almost no learning time, it is interesting for participants, and it leads to relatively animated and unconstrained conversation. Although quantitative measures of conference performance were difficult to define within the "Consensus" context, the scenario proved to be very useful for the collection of impressionistic data of the type that was of primary concern in this phase. A discussion and an example of the task are included in Appendix B.

After the two short scenarios were developed and given preliminary evaluation in the test bed, the task of formulating a scenario suitable for simulation of a large military conference remained. This goal was met by designing the "Telewar" scenario, which simulates a conference concerned with routing of military vehicles in the face of attacks by insurgent forces. This scenario, also discussed in detail in Appendix B, is far easier to learn than "Car Pool" and has the advantage of partitioning participants into different conference roles. Unfortunately, it was

found to be even less efficient than "Car Pool" as a data generator and, as a result, was employed only twice during the evaluation.

Along with improvement of the scenarios and formalization of the questionnaire, came a change in attitude concerning the utility of the fine measures of conference dynamics we had attempted to define for Phase I. Earlier, it was felt that data gathered by the computer in the form of a real-time audit trail of conference transactions would occupy an important role in the analyses of various systems. To this end, a computer program had been written that successfully captured significant transactions (collisions, speaker selections by the system, etc.). A portion of one of these trails, along with categorizations of various events that appear in the trail, is presented in Appendix D.

The collection of atomic events did not prove as useful as expected for analyses, although it was of practical value for verifying accurate functioning of the systems and, occasionally, for explicating comments made by participants during the debriefing. In view of this, we elected not to continue routine compilations of the audit trails begun in Phase I, and resolved to base the remainder of the evaluation on the more task-oriented data proceeding from ratings and scenario performance.

5.1 Statement of Purpose

During Phase II, a large number of studies were arranged among control-signal-switched, voice-controlled, and push-to-talk systems. These involved, in most instances, both simplex broadcast and speaker/interrupter protocols. The effects of delay, tandeming, and priority were investigated for selected combinations of system and protocol, and the value of various functional aids to pursuit of the chairperson role were studied.

In addition to continued study of the centrally controlled conference environment of Phase I, we began evaluation of conferences employing conditions appropriate to a distributed-control environment. The latter involved comparisons of a minimally constrained version of the broadcast-interrupter protocol discussed in Section 2.0 with distributed-control versions of the simplex broadcast and speaker/interrupter protocols of Phase I.

Finally, a limited amount of experimentation using the Telewar scenario described above was conducted with 20-person conferences using voice-controlled simplex broadcast speaker/interrupter protocols in a centrally controlled environment.

Our primary purpose throughout this phase was, of course, to establish a ranking of conditions based on measures of conference performance and participant ratings. Perhaps more *significantly, we hoped to partition the space containing the set of conditions into regions of* what might be called "relative acceptability." Our expectation was that if one could identify partitions containing conditions similar in respect to type of control, protocol or procedural constraint (delay, tandem, priority) exercised, he might then be able to infer the relative acceptability of additional conditions not currently subject to evaluation.

All eight-person conferences in this phase were conducted with the Consensus scenario, which provided a relatively realistic problem-solving context and, as explained below, enabled estimates to be made of the ease with which a chairperson could perform the task of directing the conference.

5.2 Summary of Procedure

Each of the conditions identified with this phase was categorized on the basis of its control mechanism, and all members of a given category beginning with CSS, continuing with VC and

PTT, and ending with SCDC, were run as a group. Table 3-2 of Section 3.5, reprinted here as Table 5-1, provides a schedule for the set of conditions studied.*

5.2.1 Subjects

Twenty-two 8-participant and two 20-participant conferences were conducted during the Phase II evaluation. Because of limitations imposed by regular work schedules, it was impossible to maintain the same group of eight subjects throughout the smaller conferences. However, an attempt was made to control variation among groups by choosing the 8 participants required for each conference from among a subset of 11 equally trained and experienced subjects. The two 20-participant conferences were conducted with the same set of 20 subjects.

5.2.2 Administration of Conditions

Table 5-1 presents a guide to the chronological order in which Phase II experiments were conducted. In most instances, a given experimental session included evaluation of two system conditions itemized in the table. Thus, Day 1 involved study of CSS/SB with "no delay," "no tandem," "no priority," and "chair aids," and of a similar CSS/SB system without "chair aids." Day 2 involved study of CSS/SB with "delay" and "chair aids," and a similar CSS/SB system without "chair aids," etc. Exceptions to this "two-at-a-time" rule occurred when one of the systems scheduled for experimentation did not operate properly, and, as a result, only one could be evaluated.

Each of the hour-long experimental sessions was preceded by a short briefing. This briefing included descriptions of the systems to be studied on that day and any special instructions concerning performance of the scenarios and/or completion of post-session questionnaires. When the briefing was complete, participants were released to their telephones to begin the dial-up procedure.

As indicated above, all conferences conducted in Phase II utilized a chairperson as a procedural control element. This person performed a number of functions associated with conference management, including ensuring that all participants had dialed in successfully and were in communication with each other, that scenario tasks were undertaken at the proper times, and that participants filled out their response sheets in the proper order. [An aid was also provided to participants and appears as Exhibit 5-1 (Sec. 5.2.3). As an aid to the performance of chair functions, a script, which is presented as Exhibit 5-2 (Sec. 5.2.3), was provided to the chairperson.]

Two scenarios were run on each of the systems evaluated in this phase. The scenario conducted first in each instance, Word-Go-'Round (WGR), served primarily as a "warmup" task for the participants and as an aid for troubleshooting of conference connections. The second scenario, "Consensus" for 8-participant conferences and "Telewar" for 20-participant conferences, was conducted immediately after the completion by participants of the first portion of a questionnaire identified as Exhibit 5-1 (Sec. 5.2.3).

When the second scenario had been completed and the remaining portion of Exhibit 5-1 completed, the next system was brought up and the next session consisting of WGR and Consensus (or Telewar) began. When this scenario and the relevant portion of Exhibit 5-1 had been completed, participants returned to a conference room where an informal debriefing was

*Twenty-participant conferences using VC/SB and VC/SI were actually conducted immediately prior to the beginning of PTT evaluation.

TABLE 5-1

CONDITIONS EVALUATED DURING PHASE II

System	Delay? No	Delay? Yes	Tandem? No	Tandem? Yes	Priority? No	Priority? Yes	Chair Aids? No	Chair Aids? Yes
CSS/SB	X		X		X			X
CSS/SB	X		X		X		X	
CSS/SB		X	X		X			X
CSS/SB		X	X		X		X	
VC/SB	X		X		X			
VC/SI	X		X		X (SI to chair only)			
VC/SI	X		X		X (SI to all)			
VC/SB		X	X		X			
VC/SI		X	X		X (SI to chair only)			
VC/SI	X		X		X (SI to chair only)			
VC/SB	X			X	X			
VC/SB	X		X			X		
VC/SB		X	X			X		
VC/SB	X			X		X		
PTT/SB	X		X		X			
PTT/SB		X	X		X			
PTT/SB	X		X			X		
PTT/SB		X	X			X		
PTT/SI	X		X		X (SI to chair only)			
PTT/SI	X		X		X (SI to all)			
*SCDC/BI	X		X		X			
*SCDC/SI	X		X		X			
VC/SB (20)	X		X		X			
VC/SI (20)	X		X		X			

***SCDC conferences utilized distributed-control procedures. All other conferences utilized central-control procedures. (See Sec. 2 for explanation of differences.)**

TABLE 5-2 SUMMARY OF PROCEDURES USED DURING CONDUCT OF PHASE II EXPERIMENTS		
Step	Task	Approximate Duration (min.)
1	Briefing	10
2	Dial-up and verify performance of CSS/SB (version 1)	5
3	Complete "WGR"	3
4	Complete "WGR" portion of questionnaire	2
5	Complete "Consensus"	8
6	Complete "Consensus" portion of questionnaire	2
7	Bring up CSS/SB (version 2)	2
8	Dial-up and verify performance of CSS/SB (version 2)	5
9	Complete "WGR"	3
10	Complete "WGR" portion of questionnaire	2
11	Complete "Consensus"	8
12	Complete "Consensus" portion of questionnaire	2
13	Debriefing	10

conducted. For the example cited above, then, the components of a single evaluation session were as shown in Table 5-2.

5.2.3 Discussion of Questionnaire

As indicated earlier, the data of primary interest during Phases II, III, and IV were those resulting from samplings of participants' attitudes and judgments during and after exercise of each teleconferencing system. Because of the importance attached to these data, it is appropriate to review briefly the rationale for the set of questions asked and the organization of the questionnaires that elicited the responses.

A questionnaire was designed to yield responses on four essentially different dimensions: (1) perceived difficulty of problem, (2) nature and quality of voices heard, (3) amount of effort required to use a system, and (4) perceived relative "goodness" (or "badness") of the system. The first of these was considered to be important for assessment of potential interactions between the demands of a given scenario and the perceived responsiveness of a given system in instances in which no objective measures of the former could be defined. The second was also thought to be important for assessing interactions and, in addition, for characterizing and troubleshooting possible system malfunctions. It was expected that answers to questions related to (3) and (4) would aid directly in establishing a figure of relative merit for each system. Summaries of the responses obtained with respect to item (4), relative system "goodness,"

form the bulk of the results presented in this and later sections of the report, while responses on dimensions (1), (2), and (3) provide the bases for discussions contained in Section 2.

Organization and Content of Questionnaire

The questionnaire developed for Phase II and used, with minor modifications, throughout the remainder of the year is presented as Exhibit 5-1. The format serves two purposes: (1) As indicated earlier, it guides the participant through the experimental session by directing him/her to "Read problem," "Check how voices sound," etc. (2) It elicits the desired information in a relatively straightforward manner.

Several aspects of the organization and content of the questionnaire deserve special attention. First, it will be recalled from our earlier discussion of procedure that two scenario tasks, WGR and "Consensus," were employed as conference tasks during Phase II. The word "problem" on the Exhibit refers to "Consensus," and the participant is requested to make a judgment concerning the expected difficulty of this scenario after examining it and in advance of any problem-solving experience on the system. The conference then proceeds to solve WGR and, later, the Consensus problem. Our purpose in constructing the test protocol in this way was to secure estimates on the task difficulty dimension cited above.

Second, there are two sections of the questionnaire with a checklist format. These were intended to elicit information with respect to the second dimension (nature and quality of voices) and were filled out, as required, during problem solving. The goal was to capture impressionistic data as soon as possible without interfering significantly with conference participation.

Third, the protocol requires an advance estimate of the difficulty that will be experienced while using the system to solve the (main) "problem." This estimate was expected to be based upon the limited experience gained with a system as a result of prior solution of WGR. In addition to providing further clarification of the overall system rating, it was hoped that this question, and its placement within the protocol, would aid in assessment of WGR as an evaluation tool.

Finally, the questionnaire contains a battery of items to be completed following solution of the second scenario. Most of these items resulted from review of statements made by our subjects during the informal briefings of Phase I, and, at an acknowledged risk of ambiguity during later analysis, are expressed in terms that were "natural" to them. Our intention here, as earlier, was to capture, as faithfully and as quickly as possible, impressions developed during problem solution.

A slightly longer version of Exhibit 5-1 was developed for use by a chairperson in an effort to retrieve impressions unique to exercise of the control function. A copy of this form is presented as Exhibit 5-2. (Page 2 of Exhibit 5-1 is also used by chairpersons, but is not shown in Exhibit 5-2.)

At the beginning of Phase II, participants filled out the questionnaires and then surrendered them at the end of a session. As the evaluation series proceeded and software supporting a touch-tone telephone polling function became available, the procedure was modified in such a way that participants could make their responses directly into a computer file for later analysis. Information relating to the design of this capability is presented in Section 2.

EXHIBIT 5-1

Participant Questionnaire

LAST NAME____________________ ROOM #_____ EXT_____ DIAL______ RUN #______

- - - *Dial-up*

- - - *Read problem*

```
                  '  hard   '  average '  easy   '
This problem   ---+---+---+---+---+---+---+---+---+---   to solve.
will be
```

- - - *Do word-go-round. Check how voices sound.*

____fuzzy	____clicky	____cutoff	____muffled
____nasal	____garbled	____monotonic	____squeaky
____unintelligible	____unreal	____produced by machine	

```
                  '  easy   '  average '  hard   '
Working on this ---+---+---+---+---+---+---+---+---+---
system will be
```

- - - *Do problem. Check how voices sound.*

____fuzzy	____clicky	____cutoff	____muffled
____nasal	____garbled	____monotonic	____squeaky
____unintelligible	____unreal	____produced by machine	

What was easiest about that problem?

____

____

What was hardest about that problem?

____

____

- - - *Turn Page*

```
                  '  bad    '  average '  good   '
Overall, this   ---+---+---+---+---+---+---+---+---+---
system was
```

EXHIBIT 5-1 (Continued)

This room is	quiet	normal	noisy	to work in.
Speech was	easy	normal	difficult	to understand.
I had	insufficient	sufficient	ample	time to speak.
The system produced	few	some	many	spurious sounds.
There were	many	some	few	repeat requests.
The handset, buttons, etc.were	hard	normal	easy	to manipulate.
People talked	rarely	sometimes	often	at once.
I had to speak	softer	same	louder	than usual.
My contribution was	great	good	poor	to this problem.
This system requires	little	usual	much	effort to use.
This system changed	many	some	few	voices.
This system was better than	few	some	many	other systems.
My speech was	often	usually	rarely	understood.
This problem was	dull	average	interesting	
Group performance was	poor	good	great	for this problem.
Communication is	better	same	worse	than free-air.
Work in this problem was	helped	unaffected	hindered	by the handset, buttons, etc.
I missed	many	some	few	words.
We had	little	enough	plenty	time for the problem.

EXHIBIT 5-2

Chairperson Questionnaire

LAST NAME_______________________ ROOM #_____ EXT._____ DIAL_____ RUN #_____

- - - *Dial-up and call roll*

	Conferees	Contribution	Summary of Majority Position	In Favor
1.	____________	______		______
2.	____________	______		______
3.	____________	______		______
4.	____________	______		______
5.	____________	______		______
6.	____________	______		______
7.	____________	______		______
			TOTAL:	______

- - - *Read problem #___ and instruct conferees to complete rating.*

This problem will be ---+---+---+---+---+---+---+---+---+---+--- to solve.

easy · average · hard

* 1 2 3 4 5 6 7 8 9 #

- - - *Give a "Start" signal for word-go-round #____. Check how voices sound.*

____fuzzy ____clicky ____cutoff ____muffled

____nasal ____garbled ____monotonic ____squeaky

____unintelligible ____unreal ____produced by machine

- - - *Instruct conferees to complete rating.*

Working on this system will be ---+---+---+---+---+---+---+---+---+---+---

easy · average · hard

* 1 2 3 4 5 6 7 8 9 #

- - - *Give a "Start" signal for Discussion Problem.*

- - - *Discuss your initial position and then open discussion.*

- - - *Check how voices sound.*

____fuzzy ____clicky ____cutoff ____muffled

____nasal ____garbled ____monotonic ____squeaky

____unintelligible ____unreal ____produced by machine

- - - *Summarize majority position as it emerges (column 3 above).*

- - - *Ensure that each participant has contributed. (Mark column 2 above.)*

- - - *Announce two-minute warning.*

- - - *Announce end of discussion (seven minutes from start).*

- - - *Summarize majority position and check for accuracy. Modify if necessary.*

- - - *Poll conferees to determine number in favor of position as summarized. (Mark above.)*

- - - *Instruct conferees to complete ratings and then to monitor their phones while the next system is being brought up.*

EXHIBIT 5-2 (Continued)

```
I had               |  little  |  some    |  much    |   difficulty
                 ---+---+---+---+---+---+---+---+---+---+---
                    *   1   2   3   4   5   6   7   8   9   #
  interrupting the conference to announce the two-minute warning
  and the end of discussion.

There were          |   few    | several  |  many    |   other
                 ---+---+---+---+---+---+---+---+---+---+---
                    *   1   2   3   4   5   6   7   8   9   #
  occasions on which I attempted to interrupt the discussion.

On those occasions  |  little  |  some    |  much    |   difficulty
I found          ---+---+---+---+---+---+---+---+---+---+--- gaining the floor.
                    *   1   2   3   4   5   6   7   8   9   #

I had               |  little  |  some    |  much    |   difficulty
                 ---+---+---+---+---+---+---+---+---+---+---
                    *   1   2   3   4   5   6   7   8   9   #
  ensuring that each participant contributed to the discussion.

I had               |  more    |  same    |  less    |   control
                 ---+---+---+---+---+---+---+---+---+---+---
                    *   1   2   3   4   5   6   7   8   9   #
  over the conference with this system than I would in
    a face-to-face conference.
```

5.2.4 Methods of Analysis

It became clear early in the evaluation that reasonably high levels of agreement were being obtained in regard to the overall ratings of system "goodness." As a result, our interest became focused almost completely on results obtained with this item and with that subset of other items that might aid our understanding of the reasons for the overall ratings.

Our approach to processing Phase II questionnaire data involved essentially four procedures.

(1) Adjustment of raw ratings. As might be anticipated, different participants tended to use different portions of the rating scale for a given questionnaire item over the course of the evaluation. For example, one participant might have distributed all his/her ratings between "4" and "8" on the scale, while another employed a range from "2" to "10." A third might have used a very restricted portion, say, "7" to "10."

Our assumption was that the difference in actual ranges used during the evaluation was less a matter of disagreement over the absolute "goodness" or "badness" of each of the systems rated than of differences among participants in the facility with which impressions could be distributed along the rating scale. Therefore, at what we perceived to be a very small risk of loss of absolute scale information, we concentrated our attention on deviations of a participant's rating from the mean of the ratings actually made.

For purposes of analysis and presentation in this report, the mean of each participant's ratings for a given questionnaire item over all systems, was first derived,

$$\bar{r}_{ij} = \frac{\sum_{i=1}^{n} r_{ij}}{n} \tag{1}$$

where

$\overline{r}_{ij}$ = the mean of the set of system ratings (r_i) for participant j

r_{ij} = the rating on the i^{th} system for participant j

n = the number of conferences in which participant j served.

The deviation of each participant's rating from his/her mean was then computed,

$$d_{ij} = r_{ij} - \overline{r}_{ij} \tag{2}$$

where

d_{ij} = the deviation of the rating on the i^{th} system for participant j from the mean of his/her ratings, $\overline{r}_{ij}$, over all systems.

Now,

$$\overline{d}_i = \frac{\sum_{j=1}^{m} d_{ij}}{m} \tag{3a}$$

is the mean of the set of mean deviations of ratings performed by all participants {j} for condition i, and

$$\overline{d}_j = \frac{\sum_{i=1}^{n} d_{ij}}{n} \tag{3b}$$

is the mean of the set of mean deviations computed over all conditions {i}.

The difference (Δ_p) between these two was taken,

$$\Delta_p = \overline{d}_j - \overline{d}_i \tag{4}$$

for each of the i conditions and this was the value used in subsequent analyses.

Similar operations were performed on chairperson ratings of eight-person conferences in an effort to summarize the attitudes of these participants toward the various conditions. Our intention here was to characterize the rating assigned by each chairperson to the particular condition in which he/she served as a deviation from the mean of the ratings assigned by chairpersons across all conditions. To accomplish this, we abstracted from the set of approximately 216 d_{ij}'s computed in (2) above for each questionnaire item, the subset of 22 values associated with chairpersons utilized in the 8-person conferences. The difference (Δ_c) between each of these and the mean of the set of mean deviations attributable to chairpersons ($\overline{d}_j^*$) was then computed for each item; thus

$$\Delta_c = d_{ij} - \overline{d}_j^* \; . \tag{5}$$

(2) Statistical analysis of ratings. After the system ratings had been adjusted, they were subjected to statistical analysis. The purpose of this analysis was to determine which members of the set of mean deviations obtained for a given questionnaire item differed significantly (in a statistical sense) from each other when the variation among ratings attributable to participants was considered.

Because it had proved impossible to maintain precisely the same set of eight participants over the course of Phase II, we chose to perform this part of the analysis with the Mann-Whitney U test.

The Mann-Whitney test was applied to all possible pairs of conditions $[\binom{24}{2} = 276]$ with the individual participant deviations (d_{ij}) representing replications. All tests were evaluated without a priori specification of the direction of expected differences (i.e., were two-tailed).

(3) Test for inter-participant agreement. A preliminary effort was made to determine the extent of agreement among participants concerning the dimensions of different systems as sampled by the questionnaire. For this purpose, a Kendall coefficient of concordance (W) was computed between the ranks assigned by participants across systems in response to each questionnaire item.

The practice of computing this statistic and the one identified in item (4) immediately below was discontinued when it became necessary to alter the constituency of the conference group. Since both procedures were conducted primarily to provide the experimenters with preliminary information on general trends in the data, and since the analysis described under (2) above considers the variation in participant ratings across systems, termination of this practice was considered to be of little consequence to the conduct of the study.

(4) Test of inter-item agreement. The final analysis conducted during the series was aimed at obtaining a preliminary assessment of the extent to which different questionnaire items produced the same set of system rankings across participants.

To accomplish this purpose, the rank order information implicit in the ratings performed by each participant was extracted for each system on each questionnaire item. The distribution of mean ranks for each system with respect to each item was then computed. Finally, Spearman coefficients (r_s) were determined for each possible pair $[\binom{21}{2} = 210]$ of questionnaire items with respect to the distribution of mean ranks.

This treatment was also terminated when it became necessary to alter the constituency of the conference group. However, since the interest it served was tangential to the primary goals of the project and the approach used was superficial, at best, the termination is considered to have had little impact on the progress of the evaluation.

5.3 Results

5.3.1 Overall Ratings of Systems

The results obtained with questionnaire item No. 3 ("Overall, this system was...") for all conditions studied in Phase II are presented in Fig. 5-1. The conditions have been grouped by protocol and all data have been adjusted as described in Section 5.2.3.3 above. Entries to the right of the zero point are "better" than average [i.e., $\overline{d}_i$; see Eq. (3a), Section 5.2.4]; those to the left, "worse" than average.

Three observations that can be made after examination of this figure are that (1) among all conditions evaluated, two versions of the voice control/simplex broadcast system, n and p, received the highest mean rating; (2) the three tandem conditions, one associated with a speaker/interrupter system and two associated with simplex broadcast systems account for the lowest mean ratings; (3) control-signal-switched and voice-controlled systems with delays represent mean conditions within the distribution.

One notes a considerable degree of agreement in these data with respect to the order in which ratings of similar conditions are distributed in the various systems. Thus, in all

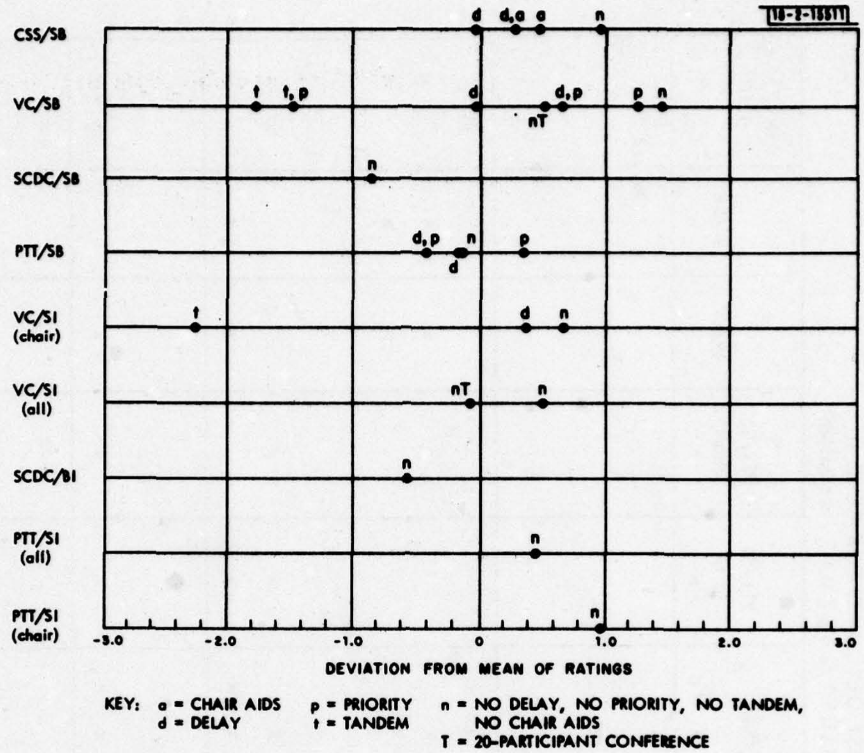


Fig. 5-1. Summary of results obtained with overall rating item during Phase II. Data have been adjusted as explained in text.

instances in which both unconstrained (n) and delay (d) conditions were evaluated [i.e., CSS/SB, VC/SB, PTT/SB, VC/SI (chair only)], the former was rated superior to the latter. Similarly, in the case of voice-control, simplex-broadcast, and speaker/interrupter (chair only) protocols, there is agreement among unconstrained (n), delay (d), and tandem (t) conditions. An exception to such agreement occurs in the reversal of n and p conditions between voice-control and push-to-talk simplex-broadcast systems.

It seems clear that increasing the number of participants to 20 (condition nT) does not have overwhelmingly deleterious effects on the ratings for either voice-controlled simplex-broadcast or voice-controlled speaker/interrupter systems under similar (unconstrained) conditions (i.e., "n" conditions). For one system in which a judgment of the relative importance of all possible treatment conditions can be judged, VC/SB, the effect of increasing conference size appears considerably less significant than effects due to tandeming (t; t, p).

The superiority of the unconstrained simplex-broadcast condition (VC/SBn) over the speaker/interrupter condition (VC/SIn) appears to be maintained in the 20-person conferences (VC/SBnT, VC/SInT), although the absolute difference between the latter pair is less than that between the former pair.

5.3.2 Statistical Analysis of Overall Ratings

Table 5-3 presents the complete set of statistically significant differences found between the points plotted in Fig. 5-1. To aid examination of the table, the order of successive rows

TABLE 5-3

SUMMARY OF MANN-WHITNEY TEST OUTCOMES FOR PHASE II RATINGS OF OVERALL SYSTEM "QUALITY" ($x = p \leqslant 0.05$, two-tailed)

			VC/SB						VC/SI (chair)		PTT/SB				PTT/SI (all)	PTT/SI (chair)	W/BI	W/SB	VC/SI (all)	
			d	t	p	P_d	P_t	n_T	d	t	n	p	d	P_d	n	n	n	n	n	n_T
CSS/SB		n		x			x			x										
		a		x			x			x										
	d	a		x			x			x										
		d		x			x			x										
VC/SB		n	x	x			x	x		x	x		x	x						
		p					x				x		x				x	x		x
	p	d		x			x										x	x		x
	n	T		x																
		d		x	x		x			x										
	p	t						x			x	x	x		x	x			x	x
		t			x	x					x	x	x	x	x		x	x	x	x
VC/SI (chair)		n		x			x			x										
		d		x			x			x										
		t			x	x		x			x	x	x	x	x	x	x		x	x
SCDC/BI		n		x															x	
SCDC/SB		n		x															x	

within a given system/protocol has been made consistent with the right-to-left (better-than-average to poorer-than-average) order of points in the figure.

In interpreting the differences presented in this table, it is important to remember that pair-wise comparisons among a large number of points may occasionally result in spurious indications of significance for a small number of the comparisons made. In an effort to reduce the possibility of spurious indications, all comparisons identified as significant in the table are based on two-tailed tests and all meet or exceed a criterion slightly more stringent than 0.05. These conventions, together with the relatively low power of the statistical test used, produce what we believe to be a more-than-adequate level of conservatism in the presentation of results.

With the aid of Table 5-3, one can make inferences concerning the actual composition of clusters of points that appear in Fig. 5-1. For example, given a particular system/protocol condition, one might determine how far to the left or right he needs to move until he encounters a condition that is reported to be significantly different. The actual point at which the critical difference is exceeded may, of course, be less than the (mean deviation) value at which the differing condition is located. In some instances, this point may be approximated by considering the significance/nonsignificance of differences associated with comparisons between the given condition and those associated with other systems/protocols. Thus, if one has established that VC/SBt,p differs from VC/SBd and wants information concerning where the cutoff actually lies, he might consider that the VC/SBd comparison with PTT/SBd is also significant and that VC/SBd-PTT/SBdp is not; therefore, the cutoff lies between the two PTT conditions, at an approximate mean deviation value of −0.3.

5.3.3 Summary of Chairperson Ratings

The overall ratings of chairpersons who served in the eight-participant conferences of this phase are presented in Fig. 5-2. Each of the points presented has been calculated in accord with the procedure presented in Section 5.2.4.

Since each datum in the figure represents, in most instances, the judgment of a single chairperson after a single run of a given condition, the distribution of ratings must be interpreted with considerable caution. Nonetheless, certain comparisons are of interest in light of results that were presented in Fig. 5-2. First, the highest rating obtained over the series was associated with an unconstrained voice-control condition (VC/SBn), while the lowest was associated with a tandem condition (VC/SBt), results which are broadly in agreement with those portrayed in the earlier figure. Moreover, tandem conditions (VC/SBt, VC/SBt,p VC/SIt) as a group represent the worst conditions encountered by chairpersons, an outcome which is also in keeping with the indications of Fig. 5-2. An analysis of results obtained on other items in the questionnaire from which these responses were obtained and on those contained in the chairperson questionnaire suggest that the reduction in voice quality, rather than a major loss in the controllability of the conference, was responsible for the ratings assigned to the tandem conditions.

Second, it is interesting to observe that both SCDC points (SI and BI) have undergone a shift to the right. The relative position of corresponding points in Figs. 5-1 and 5-2 suggests that the conferences may have been slightly more satisfactory than average from the chairpersons' points of view and slightly less satisfactory than average from the participants' points of view.

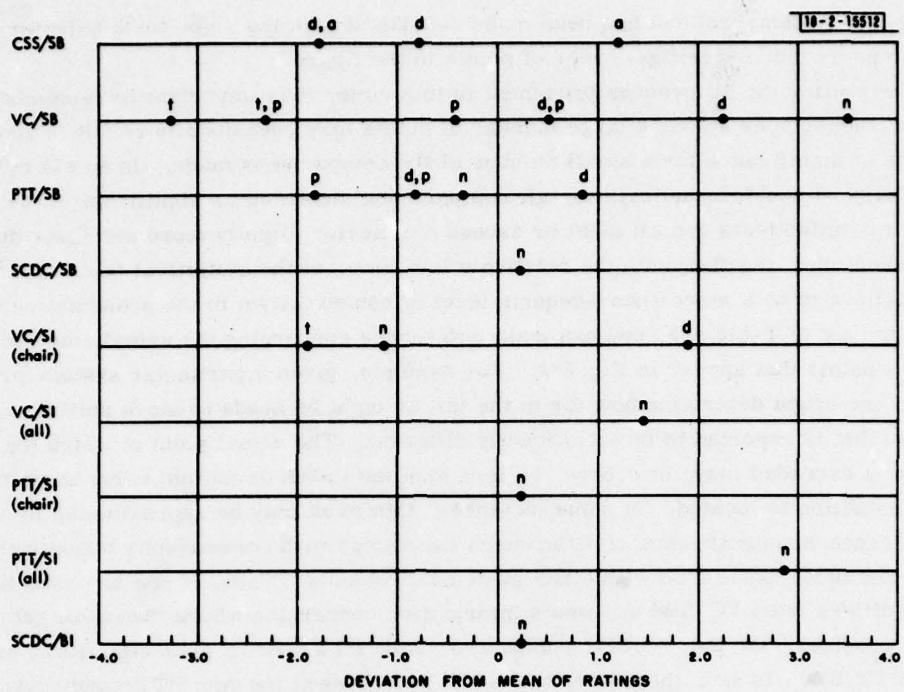


Fig. 5-2. Summary of chairperson responses to overall rating item during Phase II. Data have been adjusted as explained in text (see Fig. 5-1 for key to symbols).

Finally, it is interesting to observe that conferences conducted under delay conditions in two of the simplex-broadcast systems (VC/SBd, PTT/SBd) and in the voice-control condition (VC/SId) were judged to be far more satisfactory by chairpersons than by regular participants. It is difficult to establish the validity of these apparent differences on the basis of the small number of responses available for analysis. Other questionnaire data do suggest, however, that the slightly slower pace of conferences conducted under the delay conditions made for somewhat easier administration of chairperson duties. If this were the case, one might expect such differences to emerge. Such an hypothesis does not, of course, explain why the CSS/SBd,a and PTT/SBd,p conditions are judged to be worse by chairpersons than by regular participants.

6.0 PHASE III

6.1 Statement of Purpose

As discussed in Section 2.0, a large set of possible algorithms can be defined for dealing with major and minor collisions in distributed-control systems. Our purpose in this phase was to evaluate four alternatives from this set. In terms of the earlier discussion, these can be identified as follows:

Algorithm (A): First Speaker: Version 1. Allow first party to collision to retransmit as soon as channel becomes free.

Algorithm (B): Free-for-All: Allow all parties to collision to retransmit as soon as channel becomes free.

Algorithm (C): Random Suppression: Allow all parties to collision to retransmit with probability less-than one as soon as channel becomes free.

Algorithm (D): First Speaker: Version 2. Allow first party to collision to retransmit preamble as soon as failure to achieve crypto synchronization is detected.

In addition to investigating strategies for dealing with contention for the speech channel, it was of interest to identify the relative efficacies of different procedures for signalling the occurrence of contention. As explained in Section 2.0, the effects on rating and performance of the presence of a ("beep") signal to contenders and/or to listeners, depending upon the algorithm, was compared with the effects of no signal.

6.2 Summary of Procedure

6.2.1 Subjects

Eight subjects, four males and four females, were selected from the volunteer group on the basis of ability to serve throughout the Phase III experimentation. It proved possible to maintain the group not only through this phase but also through Phase IV.

6.2.2 Schedule of Conditions

A summary of the schedule of conditions for Phase III is presented in Table 6-1. The following conditions were replicated in order to verify their outcomes: Ab', Ad, Bd, Dd.

Conferences during this phase were conducted without benefit of chairpersons. All coordination required for starting and ending a given conference and for completing the questionnaire was handled by the experimenter via a special conference channel.

6.2.3 Method of Analysis

The only important difference between Phases II and III with respect to analysis was the selection of the more powerful Wilcoxon test for pair-wise examination of conditions. Selection of this test, which utilizes information concerning the magnitude as well as the direction of pair differences, was possible because of the availability of the same eight subjects during evaluation.

TABLE 6-1 SCHEDULE OF PHASE III CONDITIONS (Characters in parentheses are codes for Fig. 6-1)	
Collision Strategy	Signal Strategy
First Speaker: Version 1 (A)	Collider hears beep (b)
Free-for-All (B)	Collider hears nothing (b')
Random Suppression (C)	Collider hears nothing (b')
Random Suppression (C)	Collider hears beep (b)
First Speaker: Version 1 (A)	Collider hears nothing (b')
Free-for-All (B)	Collider hears beep (b)
Free-for-All (B)	Collider and listeners hear beep (d)
First Speaker: Version 2 (D)	Collider hears nothing (b')
First Speaker: Version 1 (A)	Collider and listeners hear beep (d)
First Speaker: Version 2 (D)	Collider and listeners hear beep (d)

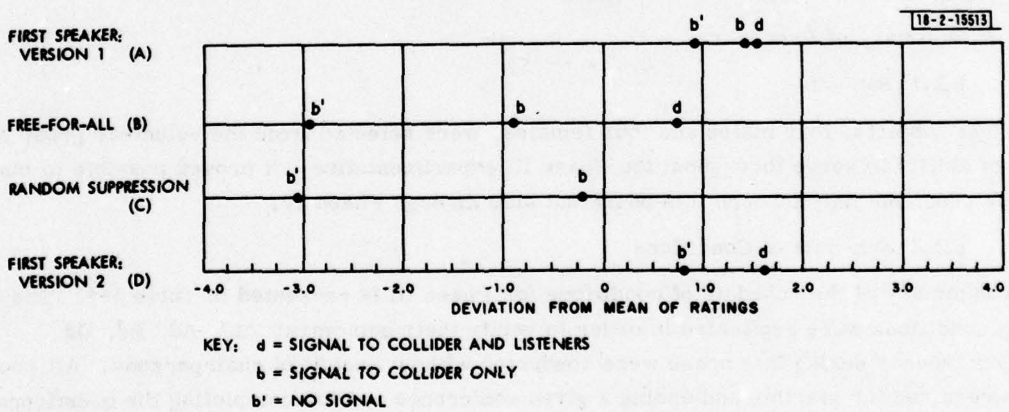


Fig. 6-1. Summary of results obtained with overall rating item during Phase III. Data have been adjusted as explained in text.

6.3 Results

6.3.1 Results Obtained with Questionnaire

The results obtained with questionnaire item No. 1 ("Overall, this system is...") are presented in Fig. 6-1. As earlier, conditions plotted to the right of zero are "better" than average; those to the left, "worse" than average.

The results portrayed here appear to be highly consistent. In all instances, conditions incorporating a collision signal, whether delivered only to the collider (b) or to the collider and listeners (d) are judged to be more satisfactory than conditions incorporating no signal (b'). Further, in the three collision strategies in which "collider only" (b) and "collider and listeners" (d) signals are compared, the latter condition always leads to higher ratings.

Finally, it seems clear from these data that A and D systems are approximately equal with respect to overall quality.

6.3.2 Statistical Analysis of Overall Ratings

A summary of results of pair-wise comparisons among points in Fig. 6-1 is presented in Table 6-2.

TABLE 6-2

SUMMARY OF WILCOXON TEST OUTCOMES ON OVERALL RATINGS FOR PHASE III

($x = p < 0.05$, two-tailed)

		A			B			C		D	
		d	b	b'	d	b	b'	b	b'	d	b'
First Speaker: Version 1 (A)	d					x	x	x	x		
	b			x			x	x	x		
	b'					x	x	x	x		
Free-for-All (B)	d					x	x		x		
	b						x		x	x	x
	b'							x		x	x
Random Suppression (C)	b								x	x	
	b'									x	x
First Speaker: Version 2 (D)	d										
	b'										

In almost all instances, judgments one is able to make concerning differences likely to be significant in Fig. 6-1 are borne out in this analysis. Thus, signalling strategies associated with A and D do not differ from each other and both systems are generally different from (superior to) B and C. In addition, "collider only" (b) conditions always prove to be superior to "no signal" (b') conditions.

6.3.3 Results Obtained with Word Match

The results obtained during this phase with the Word Match scenario are summarized in Table 6-3. It seems clear from examination of this table that performance, as measured by the percent of items correctly matched (column four), was quite good in all but one instance, Bd. However, it must be recalled that only half of the items on a given list could be successfully matched; further, that these matching items might not, as a result of their random positioning, all be examined in a given 5-min. run. A better sense of the relative level of difficulty posed by a given conferencing condition can be had by considering the number of matches actually attempted.

With "Total Attempts" as a parameter, one can detect a relatively high correlation (computed correlation, $r_s = 0.803$) of performance with the overall system ratings depicted in Fig.6-1.

TABLE 6-3

SUMMARY OF PHASE III WORD-MATCH PERFORMANCE

Collision Strategy/Signal	Total Tried	Total Correct	Percent Correct
Ab	29	29	100
Bb'	20	19	95.0
Cb'	22	21	95.4
Cb	22	22	100
Ab'	34	33	97.0
Bb	21	19	90.5
Bd	22	18	82.2
Db'	37	34	91.9
Ad	38	37	97.3
Dd	37	36	97.2

7.0 PHASE IV

7.1 Statement of Purpose

As indicated in Section 2.0, the SI protocol used in the distributed-control environment of SCDC is considerably different from that used in the centrally controlled conferences of Phase II because of the time required to switch each speaker's receiver between speaker and interrupter channels. A primary concern in this phase was to determine what, if any, effect on ratings and performance might result from variation in the length of time required for the switching process. Two values that could be expected to bracket the time actually required in an SCDC environment, 50 and 300 msec, referred to as "fast" and "slow," respectively, were studied using the First Speaker: Version 2 (D) and double-signal collision strategy of Phase III.

In addition to switching time, it was of interest to determine the effects of preamble time on ratings and performance within the SB and BI protocols of Phase III. Three different times, 24, 300, and 1067 msec, were studied using the D-double signal strategy. As indicated in Section 2.0, the shortest of these times would be expected primarily to produce "minor" collisions, while the longest would be expected to produce "major" collisions.

7.2 Summary of Procedure

7.2.1 Subjects

The eight subjects who had participated in Phase III returned to serve in Phase IV.

7.2.2 Schedule of Conditions

Table 7-1 presents the schedule of conditions for this phase. The (AB), SBS, SBL, BIS, and BIL conditions were replicated in order to verify their outcomes. Collision strategy Dd from Phase III was used throughout this part of the evaluation.

TABLE 7-1

SCHEDULE OF PHASE IV CONDITIONS

[system = SCDC/PTT, encoding = PCM, collision strategy = First Speaker: Version 2 (Dd)]

Protocol	Preamble Time (msec)	Switching Time (msec)
Delayed Analog Bridge	–	–
Simplex Broadcast (SB)	Long (L) = 300	–
Broadcast-Interrupter (BI)	Short (S) = 24	–
Broadcast-Interrupter (BI)	Extra Long (X) = 1067	–
Simplex Broadcast (SB)	Extra Long (X) = 1067	–
Simplex Broadcast (SB)	Short (S) = 24	–
Broadcast-Interrupter (BI)	Long (L) = 300	–
Speaker/Interrupter (SI)	= 24	Fast (F) = 50
Speaker/Interrupter (SI)	= 24	Slow (S) = 300

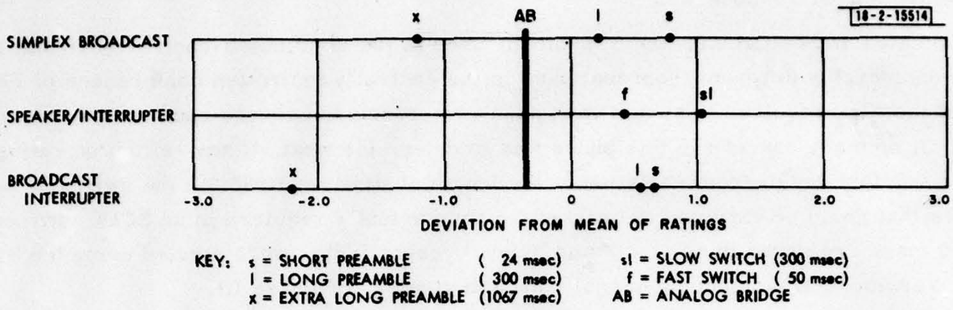


Fig.7-1. Summary of results obtained with overall rating item during Phase IV. Data have been adjusted as explained in text.

TABLE 7-2

SUMMARY OF WILCOXON TEST OUTCOMES ON OVERALL RATINGS FOR PHASE IV

($x = p < 0.05$, two-tailed)

Protocol	AB	SB			BI			SI	
Preamble/Switching Time		X	L	S	X	L	S	S	F
Analog Bridge (AB)				X	X		X	X	
SB extra long (SBX)			X			X		X	
SB long (SBL)					X				
SB short (SBS)					X				
BI extra long (BIX)						X	X	X	X
BI long (BIL)									
BI short (BIS)									
SI slow (SIS)									
SI fast (SIF)									

The major procedural difference to be noted between Phases III and IV regards the administration of Word-Match. Rather than terminate the Word-Match sessions after 5 min., as had been done earlier, the sessions were now run to completion. This permitted the conference to attempt matching each word in its collective list and provided a measure of the speed with which the total task could be accomplished. For purposes of exercising some degree of control over the total session time required for Word-Match completion, the number of items on each participant's list was reduced to five.

Conferences during this phase were conducted without benefit of chairpersons. All coordination required for starting and ending a given conference and for completing the questionnaire was handled by the experimenter via a special conference channel.

7.2.3 Methods of Analysis

Treatment of data obtained here was similar to that in Phases II and III.

7.3 Results

7.3.1 Results Obtained with Questionnaire

Figure 7-1 presents the results obtained with the various combinations of protocol and preamble time evaluated in this phase. For both SB and BI protocols, the extra long preamble (X) results in significantly lower ratings than do either of the shorter preambles. Although the latter preambles also are consistent with respect to order within protocols, it is not clear on the basis of these data that they are in fact significantly different from each other.

A surprising aspect of these ratings is that the analog bridge appears to be less satisfactory than six of the eight conditions evaluated.

7.3.2 Statistical Analysis

A summary of pair-wise comparisons of points in Fig. 7-1 is presented in Table 7-2. Here, the extra long preamble conditions (SBX, BIX) are seen to differ very significantly from the shorter ones, but between the two shorter preambles in each protocol (..L and ..S) there is no difference. No difference between switching times in the Speaker/Interrupter protocol has been demonstrated.

It is interesting that, although the conditions SBX and SBL differ significantly (0.02) from each other, the difference between SBX and SBS, which appears much greater in Fig. 7-1, is not significant. A review of the actual ratings made in these conditions suggests that this asymmetry is due to skew in both SBL and SBS distributions.

7.3.3 Results Obtained With Word Match

A summary of results obtained with the Word-Match scenario is presented in Table 7-3. Note that although differences with respect to accuracy of performance are small, there are a number of large differences with respect to task completion time. Moreover, a comparison of the order of outcomes in Fig. 7-1 with the order of outcomes here suggests a very high correlation (computed correlation, $r_s = 0.934$) between rating and performance time.

TABLE 7-3
SUMMARY OF PHASE IV
WORD-MATCH PERFORMANCE

System	Total Correct	Total Incorrect	Performance Time (sec)
Analog Bridge (AB)	39	1	305
SB extra long (SBX)	39	1	403
SB long (SBL)	40	0	285
SB short (SBS)	40	0	229
BI extra long (BIX)	39	1	368
BI long (BIL)	40	0	272
BI short (BIS)	39	1	248
SI slow (SIS)	40	0	256
SI fast (SIF)	40	0	261

REFERENCES
LINCOLN LABORATORY

1. Network Speech Processing Program Semiannual Technical Summary, Lincoln Laboratory, M.I.T. (31 March 1977), DDC AD-A045454/6.
2. Network Speech Processing Program Annual Report, Lincoln Laboratory, M.I.T. (30 September 1977), DDC AD-A053015/4.
3. Report of Studies Performed in Support of Lincoln Laboratory Program in Secure Voice Conferencing, Report No. 3681, Bolt Beranek and Newman, Inc. (March 1978).

GLOSSARY

APC	Applied Predictive Coding a speech encoding technique.
BBN	Bolt Beranek and Newman, Inc. the firm handling the human-factor aspects of the research in this report.
BI	Broadcast Interrupter an SCDC conferencing protocol (Sec. 2.7).
CRC	Communications Research Centre (Ottawa, Canada).
CSS	Control Signal Selection a conference control technique (Sec. 2.5).
CVSD	Continuously Variable Slope Delta Modulation a speech-encoding technique.
FFA	Free For All — SCDC collision-handling algorithms (Sec. 2.7).
FS-1	Favored Speaker – Version 1 — SCDC collision-handling algorithms (Sec. 2.7).
FS-2	Favored Speaker – Version 2 — SCDC collision-handling algorithms (Sec. 2.7).
IDA	Institute of Defense Analysis
LDVT	Lincoln Digital Voice Terminal a high-speed signal-processing computer developed at Lincoln Laboratory.
LPC	Linear Predictive Coding a narrowband speech-encoding technique.
ONR	Office of Naval Research
PCM	Pulse Code Modulation a wideband speech-encoding technique.
PTT	Push-To-Talk a conference control technique (Sec. 2.5).
RS	Random Suppression an SCDC collision-handling algorithm (Sec. 2.7).
SAD	Speech Activity Detector a device for determining the presence or absence of speech on a channel.
SB	Simplex Broadcast a conferencing protocol used in both centrally controlled (Sec. 2.6) and distributed-control (Sec. 2.7) systems.
SCDC	Shared Channel Distributed Control a class of conferencing techniques using distributed control of a shared communication channel (Sec. 2.7).
SI	Speaker/Interrupter a conferencing protocol used in both centrally (Sec. 2.6) and distributed-control (Sec. 2.7) systems.
VC	Voice Control a conference control technique (Sec. 2.5).

APPENDIX A

REVIEW OF SELECTED LITERATURE ON VOICE CONFERENCING

REVIEW OF SELECTED LITERATURE ON VOICE CONFERENCING

A.1 INTRODUCTION

Basic and applied studies of at least tangential interest in the context of voice teleconferencing abound. One finds in the literatures of social psychology and management many formal and informal experiments that attempt to assess the importance of leadership variables on group performance, to compare group problem solving and decision-making performance with that of individuals, to evaluate the consequences of different communication channel arrangements between group members, etc. In the literatures on speech and hearing, one finds rigorous laboratory efforts to identify critical parameters of the speech signal and to assess the effects of manipulation of these parameters on speech quality and intelligibility. The diverse literatures of human factors and artificial intelligence contain much discussion and study of human information-processing behavior and interactive problem solving.

Despite this wealth of research, our review has, to date, disclosed only a few studies that are of direct and obvious relevance to the current work. Perhaps the most significant of these from a methodological point of view is a study conducted by Richards and Swaffield in 1958. In this study, careful efforts were made to define measures of communication link performance based on the amount of effort required of users rather than on more traditional measures of information transmission. A second study of interest is that conducted by Bavelas, Orlansky, Sinaiko, and others (1963a-i) for the Institute of Defense Analysis (IDA) in the mid-sixties, for the purpose of defining and examining procedural and technical problems in telephone and teletype conferencing and for ascertaining the feasibility of conducting high-level multinational conferences. The third study, conducted recently by Chapanis and others (1972, 1974, 1975, 1977) at Johns Hopkins University, had, as its major purpose, a comparison of several different modes of communication among group members and an assessment of the fine structure of the interactive dialog.

Finally, a set of methodological studies conducted between 1939 and 1977 by Brady (1965, 1968), Jaffe and Feldstein (1970), Norwine and Murphy (1939), Phillips et al. (1977), and Williams et al. (1973) provides an important point of reference for efforts during the current project to develop computer-based methods for the analysis of speaker-interrupter dynamics.

A.2 RICHARDS AND SWAFFIELD ASSESSMENT OF SPEECH LINKS

One of the earliest and most methodologically interesting efforts to analyze the characteristics of speech communication links from a user's point of view was that of Richards and Swaffield (1958). These authors pointed out the dilemma associated with attempts to assess the quality of two-way communication links through the administration of one-way intelligibility tests:

> "Assessment over a wide range of conditions is only possible if a complete circuit is used. A complete circuit can only be achieved by considering points at or beyond A and A' [source and receiver, respectively]; this in turn requires inclusion in the link not only of the speech and hearing organs but of those brain activities associated with thinking and with idea-language transformation. Thus to make the assessment meaningful we have now had to include in the link much that is not itself part of the equipment but is part of the human user. These human parts of the circuit are liable to be of both comparatively wide variability and unknown distribution. These are circumstances that make accurate knowledge of the performance of the equipment extremely difficult to acquire.

"The only types of link that may safely be rated by one-way methods, e.g., between C and C' [sidetone paths between speech organs and ears of speakers and listeners], are thus:

"(a) Links of high performance where conversational exchange with poor talkers, listeners and conditions of use, is unnecessary to elucidate anything that is being said.

"(b) Links whose method of operation precludes return speech.

"These are relatively small classes of equipment" (p. 81).

The authors argued that complete assessment of a speech link must "in one way or another, detect (and possibly measure) the following:

"<u>For one-way conditions</u>

"1. The extent to which the reproduced speech can be distinguished from the original or what would be received over a direct air path.

"2. The use of listening (or talking) effort.

"<u>For two-way or conversational conditions</u>

"3. The use of conversational effort.

"4. Whether extra time is being taken up on account of transmission difficulties" (p. 81).

Having identified these parameters, Richards and Swaffield set up scales for characterizing a given speech link. For the two-way conversation, the "conversational effort" scale runs from "none" to "considerable," and for "message rate," from "normal" to "appreciable reduction (more than 5%)." A speech link rated with respect to these scales is then described as being "perfect," "excellent," "good," "fair," "poor or useless." The boundaries associated with each region of a given scale are defined in terms of the ratio of mean speech power (averaged over a speech period) to mean (unweighted) noise power.

The primary means exploited by the authors to estimate speaking/listening effort is opinion rating. In practice, subjects classify their judgments into one of five categories, as follows:

"A – Complete relaxation possible: no effort required.

"B – Attention necessary: no appreciable effort required.

"C – Moderate effort required.

"D – Considerable effort required.

"E – No meaning understood with any feasible effort" (p. 84).

Ratings in the various categories are rendered after subjects have had the opportunity to perform a number of tasks with a given link (e.g., the reproduction of sentences read over the link, matching of random shapes against verbal descriptions). The times taken to accomplish the tasks are also recorded, providing data for estimates of "message rate."

Quite aside from the utility of the assessment method developed, which, in the application discussed, appears significant, three aspects of the authors' thesis are of fundamental importance from a human factors point of view: (1) There is a recognition of the fact that all parties to a conversation adjust their behavior to maximize the transmission of information; thus, a "speaker" adjusts the rate and/or content of his speech to match what he perceives, on the basis of his experience as "listener," to be the constraints imposed upon a listener, while, at the same time, the listener increases his effort to acquire the message. (2) The increase or decrease in effort required of speakers and listeners represents a critical and scalable dimension of the

quality of a given speech link. (3) Ratings of speaking and listening effort may be more sensitive to manipulations of the signal-to-noise ratio of a two-way link than are objective measures of problem-solving performance.

A.3 IDA RESEARCH ON TELECONFERENCING

Eight different categories of variables were identified as important in the context of the IDA studies. Three of these are relevant to our current efforts:*

(1) Medium of Communication. A conference may be conducted on a face-to-face basis or may utilize teletype or telephone channels to connect members of the group. The relevant findings were (a) that conferences conducted by telephone were superior to those conducted on a face-to-face basis when the task was primarily one of negotiation; (b) that disparate views converge more rapidly during telephone exchanges than during teletype exchanges; (c) that tasks involving simple exchange of information are carried on more effectively over the telephone than over the teletype; and (d) that the "naturalness" of spoken language, the cues available for verification of the source of an utterance, and the feedback that can be provided in near-real time by listeners makes telephone conferencing superior to teletype conferencing.

(2) Network Configuration. Results suggested that because of the capabilities for discretionary switching, central-control networks are more conducive to the exercise of strong chairmanship than are simplex network or face-to-face arrangements. It was found, however, that four-person groups could maintain sufficient discipline in the simplex network to avoid mutual interference and the need for a chairman.

(3) Role of Chairman. Results highlighted the critical role of chairman, particularly in tasks involving negotiation, and suggested that, in conferences in which no chairman was designated, one member would emerge and fill the role as the conference proceeded. The skill required in this role was also noted, particularly when resentments grew over the control that could be exercised in a central-control network. The experimenters comment that it is important for conference participants to be familiar with the network configuration being employed and with its possible constraints, so that difficulties attributable to equipment can be clearly separated from control actions taken by the chairman.

In addition to generating an initial data base for evaluation of teleconferencing strategies and techniques, the IDA studies provide valuable insights into difficulties associated with the design of tasks and performance measures for research in conferencing. One of the initial tasks, a version of the "Traveling Salesman" game, met most of the objective criteria established by the investigators, but it was found that participants tended to engage in individual problem-solving behavior rather than to collaborate, thus minimizing the desired interaction. Though it produced the desired interaction, a second mathematical game, based on the concept of a "magic square," was found to be dull and uninteresting after a few exchanges and was eliminated from further consideration. In an effort to ensure desired levels of interaction and interest, a war game with a rich data base and a significant number of playing dimensions was developed.

* The remaining variables are (4) language, (5) staffing, (6) cultural factors, (7) channel properties, and (8) security constraints. The first three and the last of these are considered to be unique to the context in which the IDA work was prepared and not of specific interest here. No results are associated with number (7), since all experiments were conducted in a noise-free environment, though the suggestion is made that they are likely to have a significant influence in teleconferencing.

Although this approach was superior to the other two and led directly to the resource-allocation game finally employed, it proved to be too complex and to lead to irrelevant behavior on the part of the conference participants.

The point seems inescapable that trial-and-error, "development," "elaboration," and "refinement," in the words of the authors, is the only approach to successful task design in this area.

A.4 INTERACTIVE COMMUNICATION RESEARCH OF CHAPANIS et al.

The work of Chapanis et al. was similar in some respects to that accomplished in the IDA series, but it departed in significant ways from that earlier effort. First, although it was also concerned with comparative evaluation of face-to-face, voice, and written communications, the assessment of molecular activities of participants, (e.g., speaking, searching, making notes, waiting, etc.) was of far greater concern. Second, although real-world conferencing environments are of interest, the research was much less focused on a particular environment such as that which guided the concerns of IDA. Instead, it was concerned with generic aspects of interactions in whatever problem-solving environment they may occur. Third, analytic emphasis was placed on the linguistic content of queries and responses, as well as on gross measures of frequency of interaction and "tempo" of group performance.

> Problems developed for the series met six criteria:
>
> "(1) They sampled different psychological functions;
>
> (2) They were representative of tasks for which interactive computer systems were currently being used, or would be used in the future;
>
> (3) They were of recognizable and practical importance in everyday life — they were not abstract or artificial problems of the type often constructed to measure hypothetical psychological processes;
>
> (4) They had definite, recognizable solutions and the solutions could be reached within approximately an hour;
>
> (5) They required no special skills or specialized knowledge for their solution; and
>
> (6) They were formulated in such a way that their solutions required the efforts of at least two individuals working together as a team" (Chapanis et al., 1972).

Examples of problems that met these criteria and that were subsequently used successfully were a "geographic orientation problem" and an "equipment assembly problem." In the first of these, one member of the pair was required to locate either the office or home address of a physician closest to a given residence on the basis of an index of streets and a street map of Washington, D.C., in his possession, and information provided by the second member from a classified section of the telephone directory. In the second task, one member attempted to assemble an unidentified and unassembled household object (trash can carrier) on the basis of transmissions by the second member of the manufacturer's instructions.

Following is a summary of results obtained in the series of studies that are relevant to the current effort.

Influence of mode on solution time. Mean solution times associated with the voice mode were only slightly higher than those associated with the "communication-rich" ("face-to-face") mode, and were far lower than those observed with handwriting and typewriting. In experiments in which performance under combinations of modes (e.g., voice and video, handwriting and video) was examined, combinations involving voice gave rise to solution times that were significantly shorter than those associated with any other combination.

Influence of mode on allocation of activity. As might be expected, mean times associated with sending and receiving information were very similar in the communication-rich and voice modes. In addition, it was clear that in these modes, searches for parts, names, or other task materials could be carried on in parallel with the sending and receiving of information. In contrast, handwriting and typewriting modes led to serialized activity and to significantly long periods of "waiting" for the completion of messages. The basic pattern of these findings was maintained in situations in which combinations of modes were studied.

Task, job role, and mode interactions. Significant interactions were found between team-member role ("source" of information versus "seeker" of information) and problem-solving task. The authors consider that these interactions were due completely to the particular construction of the problems and to allocations of tasks between team members.

Influence of mode on verbal composition. Results of an in-depth analysis of communications between team members indicated that approximately 13 times as many words were used in modes involving voice as were used in hard-copy modes, suggesting a much higher level of redundancy and correspondingly lower information load per word in the former modes than in the latter. An effort was then made to determine what structural differences accompanied the difference in word frequency. It was found that subjects communicating in voice mode tended to use a greater number of pronouns and function words than did subjects utilizing hard-copy modes. The handwriting mode resulted in use of fewer pronouns, verbs, and verb derivatives than any other mode. These results were considered by the experimenters to be consistent with the characterization of handwriting as a telegraphic style and voice as a redundant style.

Perhaps the most appropriate observation one could make concerning the import of these results is, in the words of the authors, "The single most important decision in the design of a telecommunications link should center around the inclusion of a voice channel. In the solution of factual, real-world problems, little else seems to make a demonstrable difference" (Ochsman and Chapanis, 1974, p. 618).

Number of Conferees. A very recent study reported by Krueger (1977) in which groups of two, three, and four students engaged in face-to-face, teletype, and televoice discussions on a variety of topics of general interest, indicated that persons in larger conferences tend to use more words per message, to generate more messages, and to communicate faster.* Despite these tendencies, and contrary to the author's expectations, however, no significant differences were found among the groups of different sizes with respect to time taken to solve problems and to reach consensus. A final effect of interest was that, whereas members of the two-person conferences generated approximately the same numbers of messages, there was considerable variability among conferees with regard to message production in the larger conferences. As Krueger suggests, the tendency for domination of the conference by one or more members seemed to increase as conference size increased.

As might be expected on the basis of results obtained earlier in this series, significant differences across the three modes of communication were found with respect to measures of conference productivity. More messages and words were generated by the face-to-face groups, and communication rates were higher, in the two modes affording interaction by voice than in the

* This reference contains an excellent, comprehensive summary of work in the general area of group communication.

teletype mode. The greater difficulty associated with maintaining satisfactory interaction in the latter mode was highlighted in the comments obtained from conferees during debriefing sessions.

A.5 METHODOLOGICAL STUDIES

A.5.1 General Discussion of Technique

The growth of interest in group communication and teleconferencing has given rise to several rigorous efforts to develop methods suitable for the collection and analysis of the fine structure of conference interactions. What one generally seeks to accomplish in such an effort is to re-create, using tape recordings or computer-generated audit trails, the flow of a conversation that has occurred between conference participants. There are, however, several significant difficulties that must be overcome if the re-creation is to be a faithful copy of what one who listened to the conversation actually heard. These difficulties arise primarily out of the facts that the pattern of hesitations and pauses exhibited by a given speaker vary over time and that different speakers exhibit different articulation patterns. The problem may be further complicated by the fact that, in a computer-generated audit trail, telephone line noise may masquerade as speech unless it can somehow be identified on the basis of its temporal or spectral characteristics and then purged from the record.

The basic approach taken by investigators in this area involves two steps. In the first of these, a set of threshold values is assigned to the recorded conversation. The thresholds serve as criteria for the following: (1) rejecting energy that appears in the record but is likely to be too short to have been associated with actual speech, and (2) accepting gaps in energy that are of such a duration that they are likely to be associated with changes in articulation and normal pauses and hesitations. In addition, an estimate of the expected length of a speech "burst" may be defined. The recorded conversation is then "corrected" using the specified filling and rejection thresholds.

In the second step, an analysis aimed at accumulating information on the dynamics of conversation, is conducted with respect to the corrected record. An investigator may be concerned with the amount of time a given speaker held the floor, how often successful attempts were made to interrupt him, how much of the conference time was accounted for by speech, how conferees differed in the extent to which they contributed to the total speech, etc. The exact nature and detail of this taxonomy is generally determined by the needs of the research and differs considerably from study to study.

A.5.2 Selected Research on Methodology

The most comprehensive efforts to define thresholds suitable for the filling of gaps in speech and for rejecting spurious bursts have been conducted by Brady (1968), Jaffe and Feldstein (1970), and Phillips et al. (1972). The latter report contains an excellent summary of the work in this area and is recommended to readers interested in the methodological issues discussed in this section.

A.5.2.1 Filling Gaps and Rejecting Bursts

Of the two thresholds to be specified in what we have identified as the "first step" in the analysis, there is greater agreement concerning the threshold for filling of apparent gaps in speech. The value chosen typically ranges from 200 to 300 msec; such a value, as Phillips et al. suggests, tends to fill articulation pauses, but to leave intact hesitation and "end of clause" pauses.

Data critical to identification of this threshold are those of Goldman-Eisler (1968), who found articulation pauses to be less than 250 msec, and of Bromer (1965), who found hesitation and inter-sentence pauses to average 0.747 and 1.027 sec, respectively.

Considerably less agreement exists with respect to the specification of a threshold for rejecting bursts of energy likely due to noise artifacts. The results of Hargreaves (1960) and Norwine and Murphy (1938) indicate that units of actual speech are typically not less than 250 msec, but Brady (1965) found that approximately 20 percent are less than 200 msec. The latter investigator has employed a burst rejection threshold of 15 msec in order to include all possible speech fragments. In the most recent work by Phillips et al., a value of 300 msec is employed, but these authors recommend continuation of efforts to identify an appropriate value.

Differences among investigators also exist with respect to the basic rates at which teleconferencing lines should be sampled and their energy states ascertained, and with respect to the appropriate order in which gap-filling and burst-rejection operations should be carried out. Brady (1960) performs the sampling at 5-msec intervals and, in an effort to prevent the bridging of gaps between noise errors and speech, rejects bursts before filling gaps. Phillips et al., sample every 100 msec and fill before rejecting. Jaffe and Feldstein (1970) fill gaps implicitly by electronic means prior to sampling, sample every 300 msec, and then reject bursts as a final step.

At the moment, the implications of these differences in technique are difficult to assess, for, as Phillips et al. point out, estimated lengths of continuous speech (that is, the speech as it would appear, presumably, to a casual listener) assessed by the various methods show remarkably little variation (1.17 to 1.64 sec). However, it does seem important that research directed at understanding the implications be continued, particularly in the context of teleconferencing systems that, like those reported on here, may, by design, occasionally inhibit the free flow of conversation.

A.5.2.2 Taxonomies of Conference Events

As indicated above, the second step in the methodology involves the categorization of speech events in the corrected record of the conference. The durations and/or frequencies of events are usually accumulated with respect to certain key variables, the identity of which depends on the purpose of research. Although a complete presentation of taxonomies and definitions of all terms contained therein would be prohibitive here, appreciation for the detail that can be obtained for statistical purposes can be gained by summarizing three examples:

1. Brady (1968): Ten events are considered: (1) Talkspurt, (2) Pause, (3) Double talk (speech by two persons simultaneously), (4) Mutual silence, (5) Alternating silence (measured from the end of one speaker's talkspurt to the beginning of the other's), (6) Pause in isolation (a pause by one speaker during which another speaker is silent), (7) Solitary talkspurt (talkspurt that occurs entirely within another speaker's pause), (8) Interruption, (9) Speech after interruption, (10) Speech before interruption.

2. Jaffe and Feldstein (1970): This taxonomy includes seven events: (1) Conversation (a sequence of sounds and silence generated by two or more interacting speakers), (2) Possession of the floor, (3) Speaker Switch (a change from one speaker to another), (4) Vocalization (continuous sound

by a speaker who holds the floor), (5) Pause, (6) Switching Pause (period of pause silence by two different speakers), (7) Simultaneous speech.

3. Phillips et al. (1977): Eight categories of events are considered: (1) Floor Time (accumulated for each conferee), (2) Cycle (time measured from when a given speaker gains the floor until he gains it again), (3) Speech, (4) "Off" time (pause time and silence time of a given speaker), (5) Interruption (both "successful" and "unsuccessful" are accumulated), (6) Response (measured from end of "Speech" by one speaker to beginning of "Speech" of another), (7) Challenge (measured from time speaker gains floor until first attempted interruption), (8) Hesitancies (ratio of pause time to speech time).

The last of these taxonomies is embodied in a system called TAVI (Time Analysis of Vocal Interaction). This system, in use at the Communications Research Centre in Ottawa, provides a comprehensive analysis of voice conferences in the form of computer-generated tables and histograms and is among the most sophisticated of the methodological tools developed to date.

REFERENCES

BOLT BERANEK AND NEWMAN INC.

Bavelas, A., Teleconferencing: Background Information, IDA Research Paper P-106, Nov. 1963a.

Bavelas, A., Teleconferencing: Guidelines for Research, IDA Research Paper P-107, Nov. 1963b.

Bavelas, A., Belden, T., Glen, E., Orlansky, J., Schwartz, J., and Sinaiko, H., Teleconferencing: Summary of a Preliminary Research Project, IDA Study S-138, No. 1963c.

Belden, T., Teleconferencing: Procedures, IDA Research Paper P-111, Nov. 1963d.

Bromer, D. S., Hesitation and Grammatical Encoding. Language and Speech, 1965, 8, 148-158.

Brady, P. T., A Statistical Analysis of On-Off Patterns of Speech. Bell System Technical Journal, 1965, 44, 1-22.

Brady, P. T., A Statistical Analysis of On-Off Patterns in 16 Conversations. Bell System Technical Journal, 1968, 47, 73-91.

Chapanis, A., Ochsman, R. B., Parrish, R. N., and Weeks, G. D., Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. Human Factors, 1972, 14, 487-509.

Chapanis, A., Interactive human communication. Scientific American, 1975, 232, 36-42.

Chapanis, A., Parrish, R. N., Ochsman, R. B., and Weeks, G. D., Studies in interactive communication. II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. Human Factors, 1977, 19 (2), 101-126.

Glenn, E., Teleconferencing: Language and Cultural Factors, IDA Research Paper P-109, Nov. 1963e.

Goldman-Eisler, F., Psycholinguistics: Experiments in Spontaneous Speech. New York: Academic Press, 1968.

Jaffe, J. and Feldstein, S., Rhythms of Dialog. New York: Academic Press, 1970 (reviewed in Phillips, D. A. et al., Time Analysis of Vocal Interaction: A report on methodology. CRC Tech. Note No. 684, Ottawa, 1967).

Massachusetts Institute of Technology, Lincoln Laboratory, Network Speech Processing Program Semiannual Technical Summary; DDC AD-A045454, March 1977.

Massachusetts Institute of Technology, Lincoln Laboratory, Network Speech Processing Program Annual Report; DDC AD-A053015/4, September 1977.

Norwine, A. C. and Murphy, D. J., Characteristic Time Intervals in Telephonic Conversations. Bell System Technical Journal, 1938, 57, 281-291.

Ochsman, R. B., and Chapanis, A., The effects of ten communication modes on the behavior of teams during cooperative problem solving. International Journal of Man-Machine Studies, 1974, 6, 579-619.

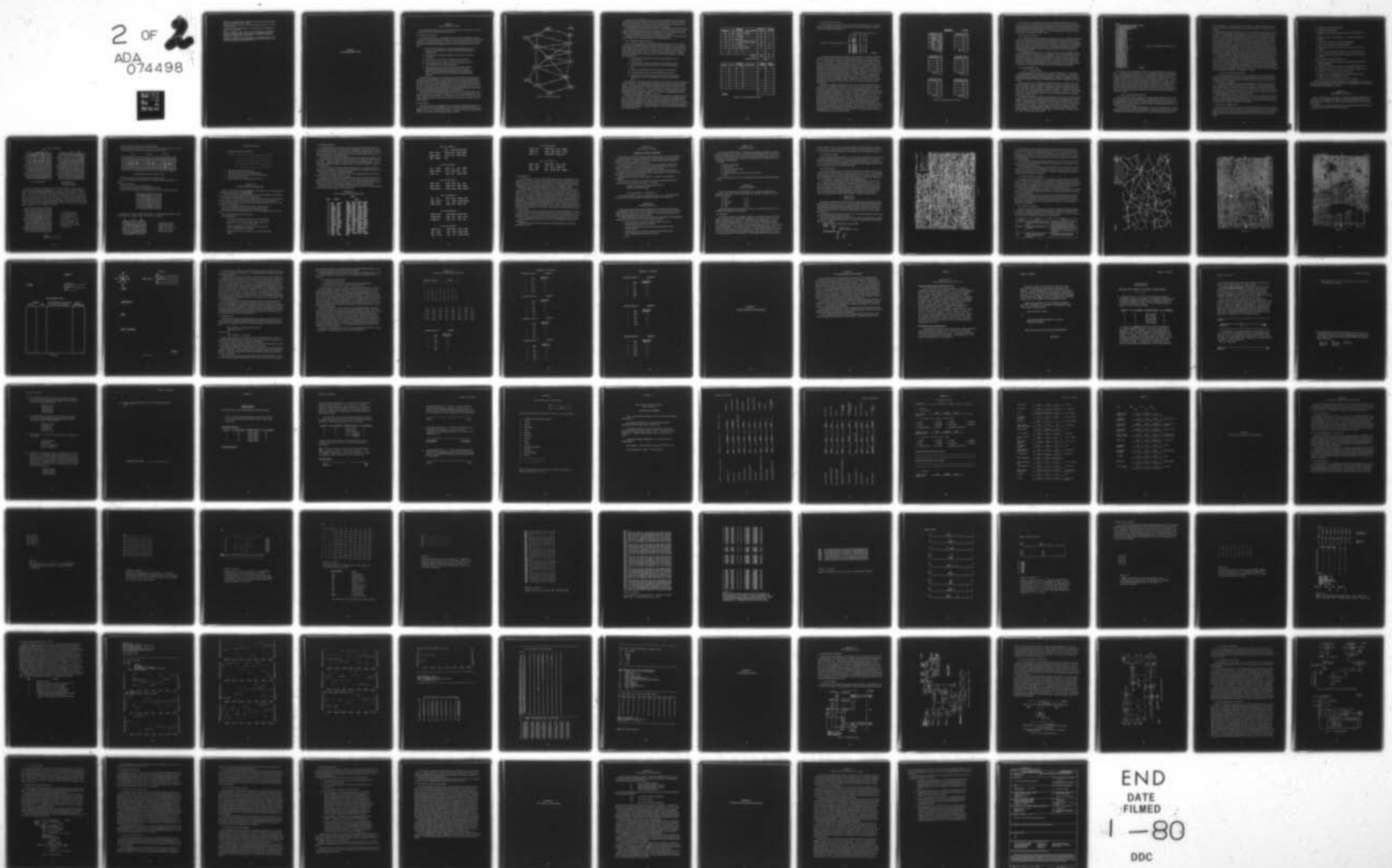
Orlansky, J., Teleconferencing: Feasibility of a Research and Development Program, IDA Research Paper P-105, Oct. 1963f.

Phillips, D. A., Treuerniet, W. C., Tigges, W. S., and Lewis, P., Time Analysis of Vocal Interaction: A Report on Methodology. CRC Technical Note No. 684, Ottawa, 1977.

AD-A074 498 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 17/2
VOICE CONFERENCING TECHNOLOGY PROGRAM, (U)
MAR 79 J.W. FORGIE, P.L. WEENE, C.E. FEEHRER F19628-78-C-0002

UNCLASSIFIED ESD-TR-79-78 N/L

2 OF 2
ADA
074498



END
DATE
FILMED
1 -80
DDC

2 OF 2

ADA 074498

Schwartz, J., Teleconferencing: Capacity of Current Communication Facilities, IDA Research Paper P-110, No. 1963g.

Sinaiko, H., Teleconferencing: Preliminary Experiments, IDA Research Paper P-108, No. 1963h.

(No Author Cited), Teleconferencing: An Experimental Task, IDA Research Paper P-112, Nov. 1963i.

Stoll, F. C., Hoecker, C. G., Krueger, G. P., and Chapanis, A., The Effects of Four Communication Modes on the Structure of Language used During Cooperative Problem-Solving. Johns Hopkins: Department of Psychology, ONR Contract No. N00014-75-C-0131, 1975.

Williams, E., Stapley, B. and Champniss, B., A Flexible System for Recording and Analyzing Speech Patterns and Other Social Interaction Events. Communications Studies Group, London, Dec. 1973.

APPENDIX B
TELECONFERENCING TASKS

APPENDIX B
TELECONFERENCING TASKS

This appendix describes the criteria for tasks, the various tasks developed and used during the project, and includes examples of the materials used.

B.1 CRITERIA TO BE MET

At the outset of this project, it was apparent that the successful evaluation of alternative approaches to teleconferencing would require the use of one or more tasks that could be employed repeatedly in a laboratory environment. On the basis of our prior experience and our reading of the reports of other teleconferencing studies, we set forth the following criteria for these tasks:

(1) The task should be usable over the entire range of conference sizes to be evaluated, and its difficulty level should be controllable independently of conference size.

(2) The task should be capable of repetition with the same set of conference participants.

(3) The task should promote continued, highly motivated performance.

(4) The task should be easy enough to learn that the participants can perform competently after a short training period.

(5) The task should permit a variety of objective performance measures, including both gross measures (such as solution time and solution quality) and fine measures of communication and system performance (such as durations of speech bursts and pauses and transitions among speakers).

Our efforts focused quickly on several tasks that required significant communication and interaction among the participants. It proved difficult, however, to find a task that met all of the criteria noted above. Several candidate tasks proved too mechanical to hold the subjects' interests. The more intellectually challenging tasks suffered from another drawback – the success of the group proved to be too dependent on the problem-solving skills of one or two of its members. The problem information was quickly shared among the participants, and then everyone proceeded to attack the problem individually, in parallel efforts.

The difficulty of finding an appropriate task has been discussed by previous workers (e.g., Aircraft Armaments, Inc., IDA Research Paper P-112, 1963), and no fully satisfactory tasks have been described in the literature. We therefore set out to develop new tasks for use in this project, using our intuitions, our collective experience, and many trial runs using the project staff as subjects.

B.2 "CAR POOL"

This section describes the development of an assignment/scheduling task that appears to have overcome some of the shortcomings of previous tasks used in teleconferencing and similar situations. The task, as it was developed, involves arranging car pools for a set of commuters, but other "cover stories" for the task could easily be employed.

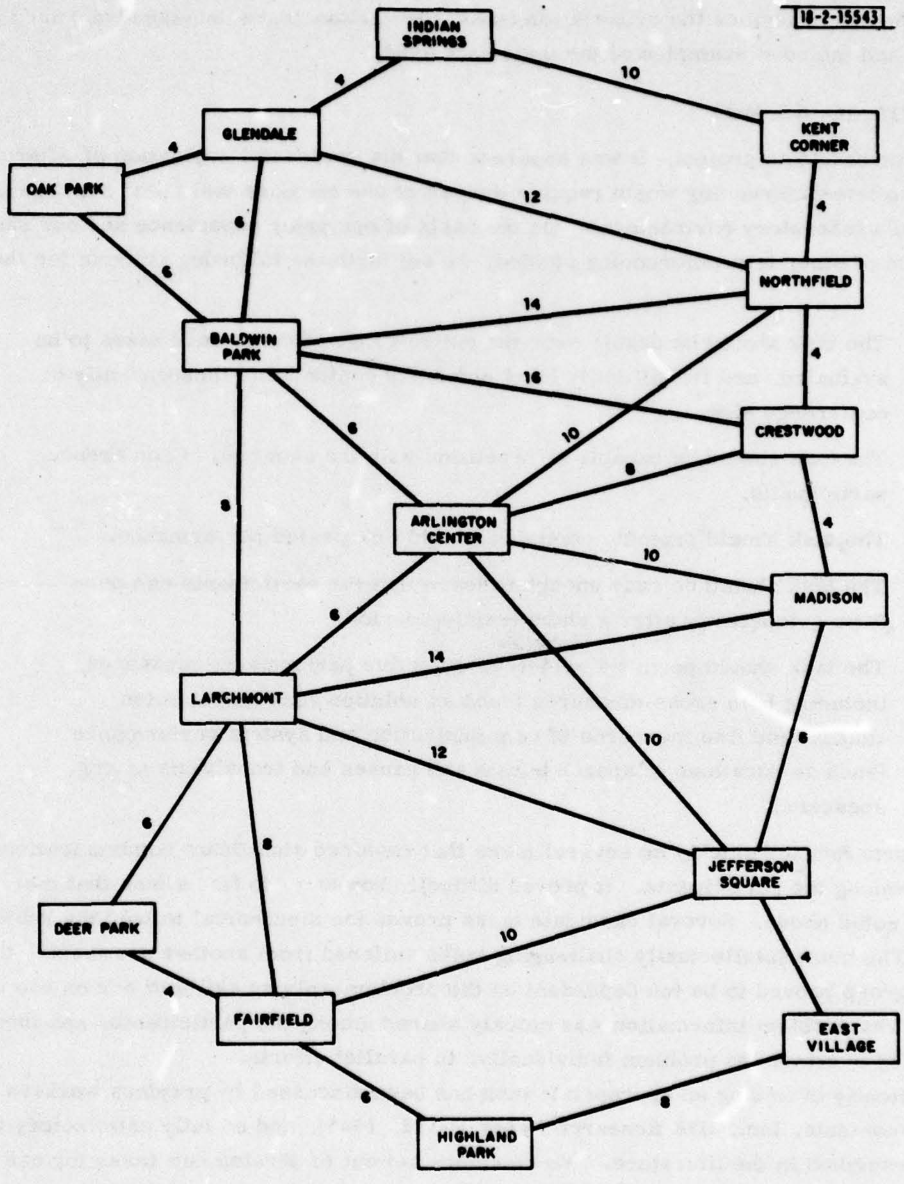


Fig. B.2.1. Standard Car Pool map.

In prior efforts at developing such a task, it proved very difficult to prevent the conference participants from simply exchanging information and then proceeding to attack the problem in parallel, individual efforts. Attempts to avoid this pitfall usually led to producing a task that was so mechanical that it failed to hold the interest of the participants.

The key elements of the task developed here were (1) a straightforward problem with simple rules, which everyone could visualize, (2) a very rich set of possible solutions, and (3) the distribution of problem-solving aids among the participants in such a way that each person found it easier to ask someone else for the result of a calculation than to perform it.

A computer program was written to aid in the generation of problems, and automatically to provide problem sheets and solution sheets for each problem generated.

B.2.1 Preliminary Versions of the Task

One of our early candidates was an assignment/scheduling task that we called "car pool." The task, as it was originally conceived, involved arranging car pools for a set of commuters in a fictitious community. Participants were given a map indicating the driving times between various points, and were told where each commuter lived and worked and what time they were required to report for work. They were then asked to assign the commuters to car pools in such a way as to minimize the total point score for all of the commuters together, under the following constraints:

(1) A commuter may arrive at work earlier than his scheduled time, but may not be late.

(2) Each commuter is assessed one point per minute of driving time and one-half point per minute that he arrives at work before the scheduled time.

(3) No points are assessed for stops to pick up and drop off commuters.

(4) No more than three commuters can be assigned to any car pool.

(5) Commuters may be picked up only at their homes and dropped off only at their offices. They are not permitted to rendezvous at intermediate points.

The map used in the preliminary version of the task was very similar to the final map, which is shown in Fig. B.2.1. The numbers shown between the "towns" on the map represent driving minutes. At the beginning of each session, each participant was given the necessary information about one or two commuters. The first step taken by the group was, of necessity, to trade information about their commuters. Then all participants set to work computing the scores of various assignment and scheduling alternatives. An example of a computation sheet employed at this stage of task development is shown in Fig. B.2.2.

It rapidly became apparent that the dominant component of the task, as originally structured, was the arithmetic computation. There were long periods of mutual silence among the participants, and communications consisted largely of comparing results and coordinating efforts to insure that no two participants were working on the same combination of commuters. The situation was clearly unsatisfactory. The task was tedious, the outcome was heavily dependent upon the problem-solving skills of the individual participants, and little communication was generated. A new approach was needed.

FROM	TO	ARRIVAL TIME	COMMUTERS	DRIVING TIME	WAITING TIME
K	N	8:26	Cook	4	–
N	C	8:30	"	4	–
C	M	8:34	"	4	–
M	J	8:40	Cook and Evans	6	–
J	F	8:50	" "	10	0
F	D	8:54	Evans	4	6
			TOTALS	32	6
			TOTAL POINTS	35	

FROM	TO	ARRIVAL TIME	COMMUTERS	DRIVING TIME	WAITING TIME

-2-18844

Fig. B.2.2. Early version of work sheet.

B.2.2 Revised Version of Task

Our first revision was designed to reduce the computational load of the task. A computer program was written to compute the component scores of various combinations of commuters and to list these scores in tabular form.

BROWN FROM GLENDALE TO EAST VILLAGE BY 830

COMS	ROUTE	DT	WT	TOT
B	GBAJE	26	0	26
BC	GBLJEJM	40	0	40
BE	GOBAJE	30	7	37
BG	GNCMJE	30	2	32
BI	GOBAJEH	38	4	42
BK	GBNCMJE	38	1	39
BCE	GOBAJEJM	40	7	47
BCG	GBCMJE	36	12	48
BCI	GOBLFHEJM	50	4	54
BCK	GBNCMJE	38	12	50
BEG	GOBACMJE	38	9	47
BEI	GOBLFHEJ	44	7	51
BEK	GOBNCMJE	42	9	51
BGI	GOGNCMJEH	46	10	56
BGK	GBNCMJE	38	4	42
BIK	GOBNCMJEH	50	7	57

-2-13843

Fig. B.2.3. Typical information sheet.

An example of the final form that was used is shown in Fig. B.2.3. This is the information sheet for commuter Brown, who lives in Glendale and must arrive at work in East Village by 8:30. This sheet contains the scores for all "reasonable" combinations of commuters in which Brown is the initial driver. For each combination, the best route is shown, along with the best score that can be achieved by optimal scheduling along this route. The points associated with driving time (DT) and waiting time (WT) are shown separately, along with their sum (TOT). The first line shows that the score for Brown driving by himself is 20 points. The next group of five lines shows the scores for all reasonable pairs of commuters in which Brown is the initial driver, and the final group shows the same information for all reasonable triplets. In each case, no entry appears for any "unreasonable" combination (i.e., one for which the best possible score is worse than that for the individual commuters driving alone). A sample work sheet (of the form finally used) for a 12-commuter problem is shown in Fig. B.2.4. The entries in the upper left-hand corner are typical of those that would be made by a conference participant during a trial solution.

The intent of this change was primarily to reduce tedium, but several other effects were observed as well. Participants now found it far easier to ask one another to look up component scores than to compute them, and a steady cross-current of inquiries quickly arose as they pursued various plausible combinations. Performance improved so rapidly that it was necessary to make the problems more difficult. It was easy, for example, to perform an exhaustive check of the approximately 100 legal solutions to a typical 6-commuter problem within 15 min. A problem of this magnitude could be solved without ever looking at the map; some of the subject groups actually tried this strategy and were successful. More difficult 8- and 12-commuter problems (with 500 to 8000 legal solutions) were employed during the experimental runs; for these problems, an exhaustive check was impossible, and every group found it essential to use the maps to focus their efforts on the more plausible candidate solutions. This meant, of course, that every problem session began with the participants calling out which commuters they had been assigned and the accompanying information about them. As this roll call proceeded, every participant annotated his or her map accordingly.

WORKSHEET -2-15540

COMMUTERS	PTS.
LAF	43
IBE	51
KG	25
JD	25
CH	27
TOTAL	171

COMMUTERS	PTS.
TOTAL	

COMMUTERS	PTS.
TOTAL	

COMMUTERS	PTS.
TOTAL	

COMMUTERS	PTS.
TOTAL	

COMMUTERS	PTS.
TOTAL	

Fig. B.2.4. Revised form of work sheet.

Other, more minor, modifications were made to the task during final shakedown, but the more crucial factors in satisfying the criteria listed above were the introduction of the tabular information sheets and the identification of the appropriate range of task difficulty to be employed. The map and problem sheet shown in Figs. B.2.1 and B.2.3 are typical of the final form of the materials given to the experimental subjects. A more detailed discussion of how the problems were generated may be found in Section 3, below.

B.2.3 Other Task Scenarios

The key elements of the revised versions of the task are (1) the use of information distributed in such a way that it is easier for the conference participants to request information from one another than to derive it themselves and (2) the establishment of the proper level of task difficulty. Clearly, there are many "cover stories" other than commuter car-pooling that could be employed. Without major changes, the point scores shown on the participant's problem sheets could have been assigned other meanings. This has not been done primarily because there has been no compelling reason to do so.

In developing other task scenarios, it is crucial to insure that there be "realistic" relationships among the problem elements and score components. If such relationships (which must almost certainly be based on a rational underlying problem structure such as the trip-scoring formula for Car Pool) are apparent to the participants, they will be motivated to employ their intuitions in formulating creative solutions; if they are not apparent, the participants will soon come to view the task as involving nothing more than a mechanical search of as many solutions as possible in the time allotted, and therefore as not very challenging or interesting.

B.2.4 Implementation of Task

B.2.4.1 Problem Generation

A computer program to generate problems was written in FORTRAN-10. This program contains standard data arrays consisting of the commuter's names, place names, and the point costs associated with traveling between adjacent places. At run time, the program requires the number of commuters to be used, their initial locations, and their destinations and scheduled arrival.

Each commuter in turn is considered as a possible initial driver. For each driver, each possible combination of one or two passengers is tried, and the route yielding the lowest point score for that pair or triplet is retained. A combination is discarded entirely if the best score found is worse than that for the individual commuters driving alone. At this point, the program outputs the number of "reasonable" pairs and triplets it has found. In Fig. B.2.5, 28 triplets and 21 pairs were retained for further analysis. These are the components out of which complete, legal solutions must be built.

The program now proceeds to combine these candidates in order to sift out and rank the legal solutions. First, however, it computes and outputs the number of combinations it will have to consider; this number provides a rough estimate of the CPU time that will be required, so that the user may abort the problem if he wishes to. In the case shown in Fig. B.2.5, 813,855 potential solutions must be checked. The vast majority of these will never actually be examined, however, because the checking algorithm is designed to eliminate the largest possible set of potential solutions for each constraint violation found. In the case shown, 5,042 legal solutions were found. The top 40 solutions are listed.

```
LIST12.6
28 BEST TRIPLETS STORED AND 21 BETTER PAIRS STORED
     813855 SOLUTIONS TO BE CHECKED
SUM OF INDIVIDUAL TRIPS IS 232 POINTS
  5042 LEGAL SOLUTIONS FOUND
171 LAF IBE KG JD CH
174 IBE DAF KG CH J L
174 LAF IBE KG DH C J
174 LAF IBE KG JD C H      [5]
174 HAC FDL IBE KG J
175 LAF ECI KBG DH J
175 LAF ECI KBG JD H
176 LAF IBE JD CH G K
176 LAF IBE KG CH D J
176 LAF KBG IE JD CH       [4]
177 IBE DAF KG C H J L
177 HAC IBE KG LF JD
177 LAF DHJ IBE KG C
178 IBE DAH KG LF C J
178 FDL IBE KG HA C J
178 ECI KBG DAF H J L
178 LAF ECI DHJ KBG        [3]
179 IBE DAF CH G J K L
179 KBG DAF IE CH J L
179 LAF IBE DH C G J K
179 LAF IBE JD C G H K
179 LAF IBE KG C D H J
179 LAF KBG IE DH C J
179 LAF KBG IE JD C H
179 ECI KBG DAH LF J
179 HAC DFJ IBE KG L
179 HAC FDL IBE G J K
179 FDL ECI KBG HA J
179 HAC FDL KBG IE J
180 IBE KG LF CH DA J
180 LAF KG DH EC IB J
180 LAF KG JD EC IB H
180 LAF EBG JD CH I K
180 FDL IBE KG CH A J
180 HAC IBE KG FD J L
180 LAF ECI KBG D H J
181 IBE KG JD CH FA L
181 IBE KG LF JD HA C
181 LAF IBE CH D G J K
181 LAF KBG DH EC I J
```

-2-18547

Fig. B.2.5. Typical problem summary sheet.

The program proceeds to generate a complete set of problem sheets, one for each commuter. A sample sheet is shown in Figure B.2.5. The problem sheet shows the best possible route (and the resulting point score) for each "reasonable" pair and triplet of commuters in which the given commuter is the initial driver. In this example, the best solution for a triplet composed of Brown, Cook, and Evans in which Brown is the initial driver carries a score of 47 points. This score must be compared with the scores that appear on the problem sheets for Cook and Evans in order to determine which commuter ought to serve as the initial driver. A particular triplet may not appear on all three sheets, of course; it may not make sense for a commuter who lives near the end of a route to serve as the initial driver, and as noted above, if the score that would result is worse than that of the individual commuters traveling alone, it will not be recorded.

B.2.4.2 Balancing Problem Difficulty

A major component of problem difficulty is the number of legal solutions that exist. This number is related, in turn, to the total number of possible solutions that must be checked, but this relationship is not a simple one. The number of possible solutions to a problem grows exponentially with the number of commuters involved (actually, the relationship involves a sum of a series of products of factorials).

In practice, however, problem difficulty appears to be a much more subtle issue, depending particularly upon how the commuters are distributed. A problem can usually be partitioned into subcomponents (for example, an eastbound set of commuters and a westbound set) that can be

attacked independently. For any given number of commuters, a "harder" problem will tend to be one in which there are more possible ways to partition the problem, each of which must be examined.

Balancing problem difficulty, then, involves several elements: (1) the number of commuters involved, (2) the total number of legal solutions found, and (3) a more subjective judgment as to the number of ways the problem can be partitioned into subcomponents. In generating specific problems, it usually took one or two iterations (but sometimes several) to produce a problem that was judged to be of similar difficulty to others in a set. It was not essential that the problems be identical in difficulty, of course; only an approximate balance was needed. The inevitable remaining variations were dealt with by means of the experimental design employed. As part of this design, pairs of problems were used which were actually identical, but in which the commuter names had been interchanged. These "permuted" problems could be presented to the same group of subjects on different days to insure a precisely even balance in problem difficulty when necessary. No problem was ever presented to the same group more than twice, and there was no indication whatsoever that any subject recognized having seen a particular problem before.

In exploring the effect of the number of conference participants on conference dynamics, it was also necessary to produce problems of equal difficulty to be solved by groups of different sizes. This presented no obstacle, because the number of commuters for which each participant is responsible could easily be adjusted. A series of eight-commuter problems was produced, to be solved by a team of eight participants (one commuter each) and a team of four participants (two commuters each). From the viewpoint of the participants, the only change in problem difficulty was that when they had two problem sheets in hand, they had to be sure they were using the proper one when responding to an inquiry. This appeared to be trivial.

B.2.4.3 *Assessing Group* Performance

As described in the body of this report, tape recordings were made of each experimental session, and the digital records made by the controlling computer were analyzed to yield micromeasures of conference dynamics. Some macromeasures could be generated, however, even as the conference was in progress.

One experimenter monitored each session using headphones. Using an audio signal on the tape, he marked the start of each session and started a stopwatch. With the solution sheet in hand, it proved fairly easy to follow the progress of the group as they tried different solutions, and the times at which these solutions were reached could be recorded. Macromeasures obtained included the total number of legal solutions generated, the times at which they were reached, the best solution obtained, and the time at which it was reached.

The bracketed numbers in Fig. B.2.5 represent the order in which certain solutions were obtained by a particular group in one experimental session. The first two solutions scored 184 and 182 points, and were not among the top 40 solutions listed. A sixth solution, scoring 190 points, was also obtained just before the conference was terminated.

B.3 "PATH" AND "NUMBER PASS"

The two tasks presented in this section were used infrequently during conduct of the research discussed in this report. They are, however, considered to be ideal for the study of systems that impose explicit constraint on the speed and accuracy with which information can be transmitted around a conference. Specific desirable characteristics of the tasks in such a context are as follows:

(1) Messages are completely determined.

(2) Message length is controlled.

(3) Messages can be short, so entries into system can be frequent in unit time.

(4) Each word, digit, or letter is critical, i.e., there is no redundancy in context.

(5) Conferees have equal or nearly equal "speaking parts."

(6) Errors are almost always immediately apparent.

(7) Task-induced errors are rare since the tasks are easy to learn and to perform.

(8) There is practically nothing to remember and no calculations to make.

(9) No task operations intervene between detecting the cue and speaking the message.

(10) It is easy to create equivalent sequences.

(11) Sequences can be reused even for the same participants since there is no gain in memorizing.

(12) The tasks are suitable for any number of participants greater than two.

(13) Deliberate errors can be inserted to induce cooperative problem solving.

(14) Visual or other "noise" can be added to either task.

(15) Either task can be constructed and/or used so that specified system qualities, e.g., voice recognition, are tested.

(16) The tasks can be used in a larger, more realistic scenario.

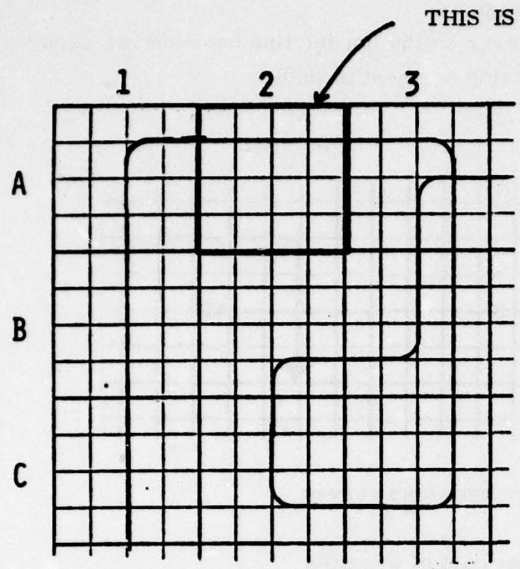
A detailed understanding of "Path" and "Number Pass" can be gained from the formal instructions which were presented to subjects during early training sessions. These instructions are reprinted below as Exhibits B.3.1 and B.3.2.

EXHIBIT B.3.1

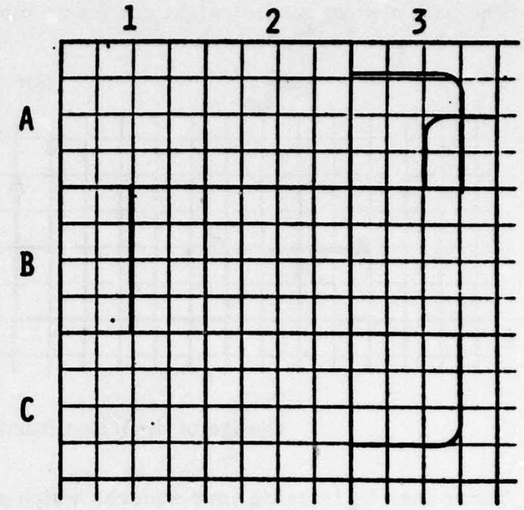
INSTRUCTIONS FOR "PATH"

"Path" is a task for two or more people. A continuous line has been drawn on a piece of graph paper. The graph paper has heavily ruled 1-in. squares and lightly ruled lines every 1/4 in. Each square is identified by a letter (along the left margin) and a numeral (along the upper margin).

Each person has a sheet of graph paper with several squares from the original, along with the sections of the continuous line drawn in those squares.

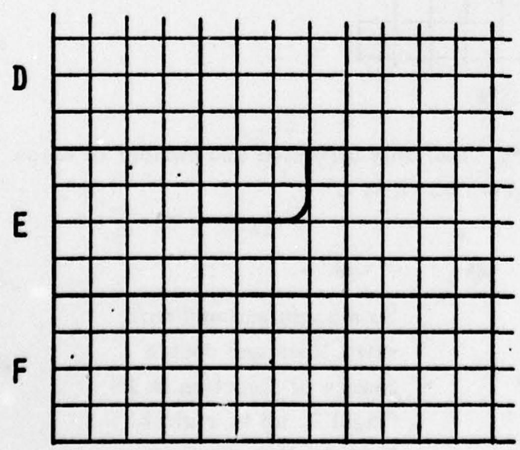


Part of the original.



This person has squares A3, B1, and C3 from the original.

The task is to follow and complete the path. We will say who starts. One person says: "Begin at edge of square (letter), (numeral), e.g., A10 or G3. Everyone will have a dot marked there. The person with that square then tells what the path does in that square, e.g., "left 3, down 1". The only words you need are "left, right, up, down" and "one, two, three, and four"; the "one, two, three, and four" refer to the 1/4-in. lines. The person with the 1st square says what is needed to get the line to an inside edge of the square. The person with the square touching that edge then takes over and says the information needed to continue the path to touch another square. The process continues until the path crosses out past a dot. The person with that square should announce the end, e.g., "Right 4, end at dot."



For example, the path may be coming down from D2. the person with square E2 says: "down 2, left 3" and the person with square E1 continues. You need not say the coordinates of your square.

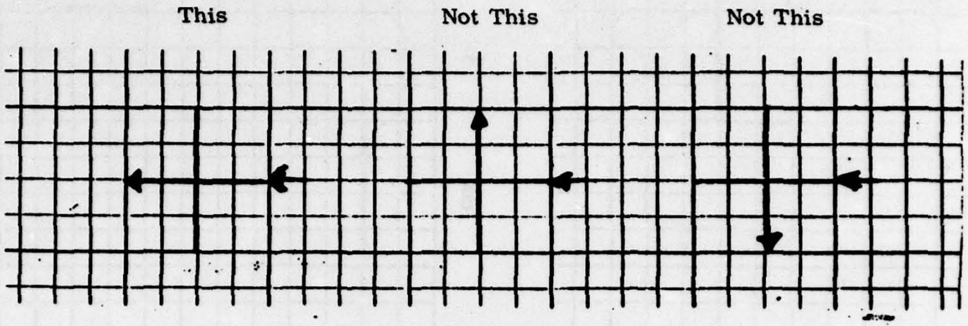
NAME ______________

DATE ______________

Certain rules have been followed in making up the paths:

The path is always on one of the 1/4-in. lines, never on the border line between two squares.

The path always goes straight across an intersecting segment of path.



change of direction (turns) are shown as curves

There may be lines on your squares which are not part of the path.

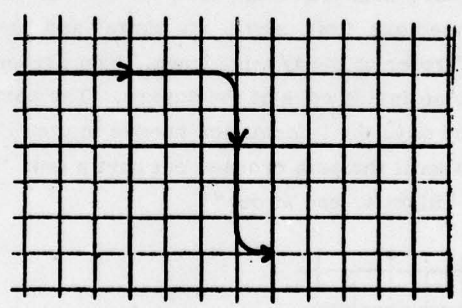
You never use the same part of the path twice. The path always begins and ends at the margin.

When you do the task:

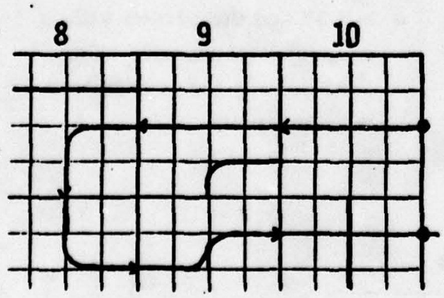
Put your name and the date in the place indicated.

Use a pencil and draw all the parts of the path you don't have.

Draw a little arrow to indicate the direction of the path, when it touches a new square.



Work as fast as you can, consistent with accuracy. Use only direction and number of steps to communicate, e.g., "right 4," "right 3, down 2," "down 3, right 1."



Note begin and end dots, extra lines and double change of direction in B9 "right 2, up 1, right 2." Note the little arrows.

Instructions for Starters

1. Check that everyone is on the phones.

   Ask for: ______________________________

   ____

   ____

   ____

   ____

2. Check that everyone can hear and speak.
3. Check that everyone has a pencil and workspace.
4. Check that everyone has put name and date on graph paper.
5. Check that everyone is ready.
6. Say the position of the starting dot and start the path.

EXHIBIT B.3.2

INSTRUCTIONS FOR "NUMBER PASS"

Number Pass is a task for three or more people. Each person uses a set of cards and each card has two sets of numbers, e.g., $\boxed{58 \quad 362714}$.

You must listen for the two digits on the left of the card you have showing. When you hear them, you say your six-digit number and turn to the next card.

It is easier to listen for the first of the two digits on the left of your card and, if you hear it, listen for the other. They must be in order and consecutive for you to say your number.

Say each of the six digits on the right of your card clearly, Say "three, six, two, ..." and not "thirty-six, twenty-seven ..."

If you say your six-digit number and no one follows, repeat the number.

The cards are arranged so the order of people speaking will change.

There is only one way to get through the pack, so you must be attentive and not miss your turn.

The task will continue until you are told to stop.

If you are the starter:

Check that each person is on the phone and can hear and speak.

Check that each person has a card pack open to the first card.

SAY: "__________, Are your cards ready?"

(If you are not the starter, make sure you can hear each other person. Tell the starter if you cannot.)

When everyone is ready, SAY "Ready, Go" and say your six-digit number.

B.4 "WORD-GO-'ROUND"

The "Number Pass" task used during training sessions and informal Phase I experiments was revised for use as a quick test of voice quality, intelligibility, and turn-around time in current teleconferencing systems. The revised task, called Word-Go-'Round (WGR), is similar in form to the earlier task, but substitutes words that may be confused and lead to critical errors when transmitted over low-bandwidth channels.

The word list used to generate the sample WGR was chosen to have low confusability. Other word lists could be used to change the level of task difficulty or to emphasize specific features of systems. A computer program was written and used to generate WGR materials. The word list, number of speakers, number of rounds per speaker, and number of words in a "call" are arbitrarily chosen.

A copy of one protocol for WGR is attached (Exhibit B.4). The first page is the experimenter's script which shows the sequence and allows progress to be monitored and timed. Successive pages are each held by individual participants.

Learning time, including a practice run, was 15 min. Running time, for the three-round, eight-speaker protocol shown, would be 2 to 3 min.

The primary output measure is task performance time, which we believe depends on ease of understanding speakers and ease of system use. Secondary measures include requests for repeats, errors made, and the nature of errors made. We expect the secondary measures to vary with noise in the system, speech quality, and attentional and other individual factors.

WGR was used prior to each "consensus" task in Phase II and served as a warmup, allowing each participant to hear all others and to use the system.

EXHIBIT B.4
Script Protocol for WORD-GO-'ROUND Task

SCRIPT

SPKR	TRIGGER		CALL			
1			HAM	FLAKES	CEREAL	STEAK
2	HAM	FLAKES	STEAK	MILK	MILK	STEAK
3	MILK	MILK	STEAK	TOAST	HAM	HAM
4	HAM	HAM	BACON	EGGS	CEREAL	TOAST
5	CEREAL	TOAST	COFFEE	HONEY	COFFEE	MILK
6	COFFEE	MILK	HONEY	EGGS	JUICE	TOAST
7	HONEY	EGGS	COFFEE	CREAM	TOAST	POTATO
8	CREAM	TOAST	COFFEE	CHEX	JUICE	CREAM
1	CHEX	JUICE	MUFFIN	HONEY	FLAKES	CREAM
5	FLAKES	CREAM	CREAM	MILK	CREAM	JUICE
6	MILK	CREAM	FRUIT	HONEY	POTATO	EGGS
3	POTATO	EGGS	POTATO	MILK	HONEY	STEAK
8	MILK	HONEY	MILK	TOAST	COFFEE	TEA
7	COFFEE	TEA	HONEY	COFFEE	SYRUP	POTATO
2	SYRUP	POTATO	JUICE	MUFFIN	CHEX	EGGS
4	JUICE	MUFFIN	TEA	HONEY	FLAKES	POTATO
1	FLAKES	POTATO	GRITS	MILK	WAFFLE	STEAK
3	WAFFLE	STEAK	HONEY	STEAK	STEAK	WAFFLE
4	STEAK	STEAK	STEAK	CHEX	BACON	TEA
6	CHEX	BACON	EGGS	HAM	POTATO	POTATO
7	EGGS	HAM	FLAKES	BACON	FLAKES	TEA
5	FLAKES	TEA	TOAST	HAM	FRUIT	HAM
2	FRUIT	HAM	HONEY	MILK	MUFFIN	SYRUP
8	HONEY	MILK	EGGS	COFFEE	CHEX	FLAKES
1	EGGS	COFFEE	DONE			

LIST FOR SPEAKER 1

		HAM	FLAKES	CEREAL	STEAK
CHEX	JUICE	MUFFIN	HONEY	FLAKES	CREAM
FLAKES	POTATO	GRITS	MILK	WAFFLE	STEAK
EGGS	COFFEE	DONE			

LIST FOR SPEAKER 2

HAM	FLAKES	STEAK	MILK	MILK	STEAK
SYRUP	POTATO	JUICE	MUFFIN	CHEX	EGGS
FRUIT	HAM	HONEY	MILK	MUFFIN	SYRUP

LIST FOR SPEAKER 3

MILK	MILK	STEAK	TOAST	HAM	HAM
POTATO	EGGS	POTATO	MILK	HONEY	STEAK
WAFFLE	STEAK	HONEY	STEAK	STEAK	WAFFLE

LIST FOR SPEAKER 4

HAM	HAM	BACON	EGGS	CEREAL	TOAST
JUICE	MUFFIN	TEA	HONEY	FLAKES	POTATO
STEAK	STEAK	STEAK	CHEX	BACON	TEA

LIST FOR SPEAKER 5

CEREAL	TOAST	COFFEE	HONEY	COFFEE	MILK
FLAKES	CREAM	CREAM	MILK	CREAM	JUICE
FLAKES	TEA	TOAST	HAM	FRUIT	HAM

LIST FOR SPEAKER 6

COFFEE	MILK	HONEY	EGGS	JUICE	TOAST
MILK	CREAM	FRUIT	HONEY	POTATO	EGGS
CHEX	BACON	EGGS	HAM	POTATO	POTATO

LIST FOR SPEAKER 7

HONEY	EGGS	COFFEE	CREAM	TOAST	POTATO
COFFEE	TEA	HONEY	COFFEE	SYRUP	POTATO
EGGS	HAM	FLAKES	BACON	FLAKES	TEA

LIST FOR SPEAKER 8

CREAM	TOAST	COFFEE	CHEX	JUICE	CREAM
MILK	HONEY	MILK	TOAST	COFFEE	TEA
HONEY	MILK	EGGS	COFFEE	CHEX	FLAKES

B.5 "CONSENSUS"

The "Consensus" task, used extensively during Phase II, requires participants to reach agreement on a course of action that represents the best response to a hypothetical problem posed at the outset of the conference. In its typical form, the statement of a given problem is accompanied by three or four alternatives that conferees are encouraged to consider, but encouragement is given by test administrators for the evolution of additional solutions.

"Consensus" has the virtues of (1) being easy to administer, (2) being intrinsically interesting as a task, (3) requiring almost no prior training, (4) being usable over a wide range of conference sizes, and (5) providing an environment in which chairperson functions, voting procedures, priority schemes, and speech collision avoidance techniques can be manipulated. Perhaps its chief disadvantage is that problem statements cannot be reused with the same group of participants. "Consensus" is an unstructured dialog task and can be a veridical simulation of actual system use. Participants gain global experience with the system in use, and their experience is tapped with rating scales and questionnaires. As such, "Consensus" has no specific output measures, although secondary measures, such as the time required to interrupt the conference and provide a message, could be made. Other measures of conversational interaction could also be made.

The problem was posed on the system well in advance of the opening statement. Conferees were allowed a total of 7 min. discussion and additional time was used by the chairperson to sum up and poll conferees. The 7-min. period was barely adequate for most conferences in terms of discussion, but was apparently sufficient for conferees to gain experience with the system.

A set of instructions was prepared to serve as a guide to participants and it is presented as Exhibit B.5.1.

Examples of "Consensus" problems employed in our experimentation are provided as Exhibits B.5.2 to B.5.4.

EXHIBIT B.5.1

Instructions for "Consensus"

COMMENTS ON "SHORT PROBLEMS"

The problems are frameworks for group conversation and creativity. You need not express your feelings, attitudes, or opinions during the short conversations to be held. You may adopt or play any role you like, as long as it is consistent, possible within the problem framework, and provides adequate opportunity for participation.

We are interested in the ease, efficiency, and quality of your conference within the constraints of the system used. We are not planning to make fine judgements on the particular solution or consensus you reach; time constraints will limit your efforts.

You may state and use any reasonable assumptions which do not contradict known facts. For example, if we were to do the moon problem again, you could assume a "day-zone" problem and rank order the items with that constraint.

Although you need not agree on a plan; negotiation, compromise, and agreement are encouraged.

At some point in the conference, you must give a clear, oral presentation including the course(s) of action to be taken and your reasons. For example:

> "We should hire Miss Smith, because she has the most experience with that product line."
>
> or
>
> "Four of us agree that ... and six think ... "

We will give time signals with 2 min. to go and with 30 sec to go. You may use the entire period as you like, but you may find the presentation best placed at or near the end.

EXHIBIT B.5.2

Consensus Problem No. 13

Recent studies of the Earth's activity in California have convinced some scientists that a major earthquake is likely to occur within the next three years. Other scientists are convinced a quake will occur but that it will be of small magnitude. A third group finds the studies completely unconvincing and believes no quake will occur.

As government officials concerned both with the safety of the population and with maintenance of the economic base of the state, what course of action will you take?

1. Fund more studies in the hope that a better prediction can be made (average length of past studies = 12 months).
2. Notify the population of the possible risk and let people choose their own courses of action.
3. Evacuate areas where the effects are expected to be worst if the quake occurs; help businesses in those areas to relocate.
4. Other.

EXHIBIT B.5.3

Consensus Problem No. 2

You are the Joint Chiefs of Staff and armed guerrilla insurgents have attacked the seaport capitol of a nonaligned nation. Many U. S. citizens work there and many more are tourists there. The nation is within easy airplane and missile range and the Third Fleet is on a training cruise 600 miles away. Intelligence sources indicate a 0.65 probability of an insurgent victory within 5 days, under present conditions.

You may:

(a) Issue a stern warning.
(b) Send the fleet.
(c) Say you are sending the fleet.
(d) Bomb the capital.
(e) Airlift troops and hold the airport clear for evacuation.
(f) Other.

The President is going to make a speech on television and requires your plan and rationale in 10 min.

EXHIBIT B.5.4

Consensus Problem No. 8

You are in the design department of PERSONS, INC. You have been asked to build a "LAWYER." (Your most successful previous project was "DOCTOR" – thousands were built). Specify the optimal mix:

(a) honesty	____%
(b) guts	____%
(c) knowledge	____%
(d) other	____%
(e) trace	5%

The prototype department needs the specs in 10 min., so you can leave 5% for trace characteristics and concentrate on important ones.

B.6 "TELEWAR"

The "Telewar" scenario was developed by BBN for use during the second phase of the Lincoln Laboratory effort in secure voice conferencing. As a tool for the study of experimental teleconferencing arrangements, it meets the criteria established for conference tasks and problems identified earlier in this report (see Sec. 3.1) and provides somewhat more freedom of choice over the selection of experimental variables than did the car-pool scenario employed during Phase I. This freedom could be of great benefit in assessing the value to chairpersons of particular conference control capabilities. A single Telewar session can take as little as 30 min. or can be used to provide a continuing problem for a number of sessions. Telewar can conveniently employ 12 to 25 participants. Telewar requires explanation to participants and

practice sessions. Very exact, but unconstrained interaction is required to produce solutions. The primary output measure could be either time to solution or quality of solution; however, these are not currently considered as important as the methods used to tap participant experiences.

B.6.1 Elements of Telewar

Telewar contains three major elements: (1) a "cover story," in the context of which resource allocation problems are defined and are solved by conference participants; (2) a set of quasi-military roles that are assumed by participants and that guide interactions over conference lines; (3) a set of procedures that must be employed by participants during a teleconferencing session. Brief summaries of each of these elements appear below:

B.6.2 The Cover Story

The scenario currently in use is concerned with a limited war that begins with an attack by "Enemy" forces on "Friendly" forces in a region of southern Germany. As the scenario unfolds, the Enemy extends its hold on the region until the Friendly forces, as a result of judicious allocation of defensive resources, is able to bring the Enemy progress to a halt. The scenario then enters a second phase, during which the Enemy is gradually pushed back.

The substance of the scenario is carried in a set of "Situation Reports" and three sets of maps. A Situation Report (see Exhibit B.6.1) contains three pieces of information: (1) a summary of Friendly and Enemy tactical activity for the simulated period just prior to the current experimental session; (2) a statement of the resource allocation objective(s) to be pursued in the current session; (3) a brief summary provided by a simulated G-2 unit concerning cities, roads, and intersections that cannot be employed during the session for the transport of resources because of sabotage, refugee traffic, flooding, or occupation by Enemy forces.

EXHIBIT B.6.1
Situation Report

At last report, Friendly units of the 105th Division had been overrun at Saalfeld and further advances made in sectors 1.3, 1.4, and 1.5. The Forward Edge of the Battle Area now extends from Saalfeld and Teuchern through an area slightly southwest of Taucha and on to Riesa. Information obtained from prisoners captured during the advance on Teuchern indicates that the next Enemy objective will be to drive due east from Saalfeld in an effort to encircle Friendly forces at Gern.

Our objectives will be to reinforce the units at Gern and at Hof with infantry and armor units of the 10th Division located at Slany and Teplice.

Many roads and intersections in the Reichenbach area continue to be blocked due to sabotage and to the flow of refugees from the northwest. G-2 reports the following roads, intersections and cities to be unusable.

CITIES: SAAFELD, TEUCHERN, TAUCHA, RIESA

ROADS:

Sector	
2.3	from #1 to #2
2.5	from #1 northwest to border

INTERSECTIONS:

Sector	
2.2	#1
2.4	#3, #4
2.5	#3
2.6	#1
3.3	#4

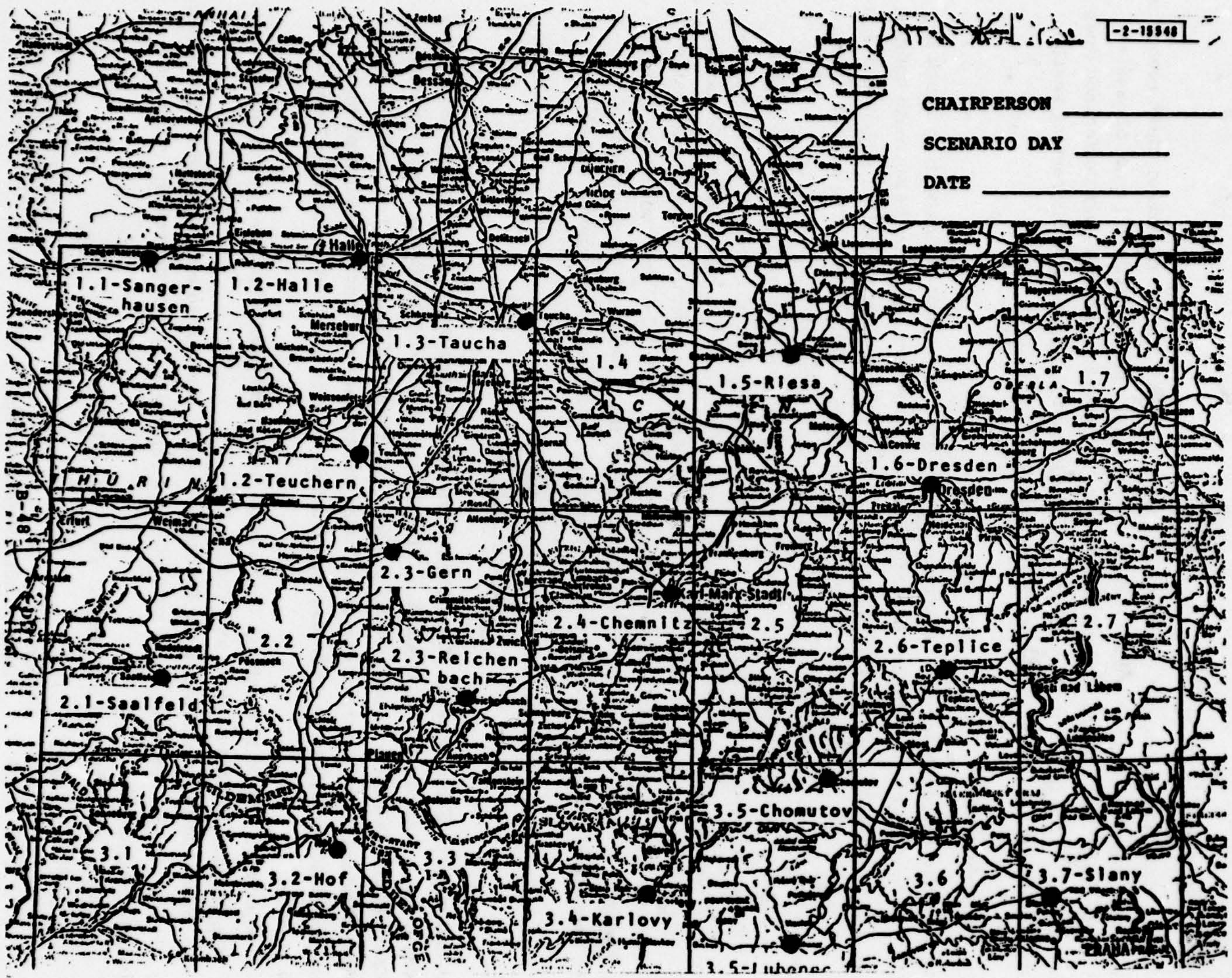

-2-15948
CHAIRPERSON
SCENARIO DAY
DATE
1.1-Sanger-
hausen
1.2-Halle
1.3-Taucha
1.4
1.5-Riesa
1.7
1.6-Dresden
1.2-Teuchern
2.3-Gern
2.4-Chemnitz
2.5
2.6-Teplice
2.7
2.2
2.3-Reichen-
bach
2.1-Saalfeld
3.1
3.2-Hof
3.3
3.4-Karlovy
3.5-Chomutov
3.5-Lubenec
3.6
3.7-Slany
B-48

Exhibit B.6.2

The first member of the set of maps employed is a reproduction of the area of Germany lying between 11° and 14°30' E. longitude and between 51°30' and 50°30' N. latitude (see Exhibit B.6.2). On this map, each 30' of latitude and longitude are intensified in order to provide a prominent grid. Each 30' × 30' sector is provided with a sector number and, in most cases, each also contains one or two large dots that mark the locations of cities that play a prominent role in the scenario.

The second member of the set of maps (see Exhibit B.6.3) is a reduced form of the first in which only key cities and the roads connecting them are represented. The grid referred to earlier is not presented on these maps.

The final set of maps (see Exhibit B.6.4) contains enlarged views of each of the sectors portrayed in the grid map. The actual relationships of the roads to the cities they serve are more evident here, and each intersection is numbered. The maps are reproduced with a simulated terrain overlay for purposes of realism only.

B.6.3 Roles of Conference Participants

Three roles are specified for Telewar: (1) chairperson, (2) tactical planner, (3) staff support. The chairperson is responsible for reading the Situation Report at the outset of a teleconferencing session, and, as explained in the section below, for coordinating dialog between tactical planners and staff. The chairperson is also responsible for preparing a summary of the route structures developed by planners over the course of the session (see Exhibit B.6.5).

With the aid of the maps depicting the complete set of cities and interconnections, the tactical planners are responsible for choosing routes that form uninterrupted paths between cities identified in the Situation Report.

Participants assuming the role of staff support use the individual sector maps to supply detailed information concerning the status (usable/unusable) of roads and intersections within a given region. This information is accumulated at the beginning of a session as a result of monitoring the the intelligence portion of the Situation Report and is noted on a "Damage Report" for use later in the session.

B.6.4 Conference Procedures

In order to understand the basic problem to be solved by communication during the playing of Telewar, it is necessary to recognize what information is available and what information is not available to participants assuming each of the roles specified above. A summary of the contents of each of these categories is presented in the following table.

Role	Known	Unknown
Chairperson	relationship between sectors and cities and between sectors and some prominent roads	complete set of roads nominally available; usability of particular roads over entire area; weightings assigned to intersections
Planners	complete set of roads nominally available for use	precise relationships between sectors, road segments, and cities; usability of particular roads within sectors; weightings assigned to intersections
Staff	usability of particular roads within sectors; weightings assigned to intersections	relationships between sectors and cities and between sectors and road segments

-2-15549

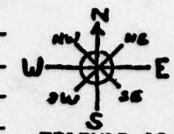
NAME ____________

CALL SIGN ____________

SCENARIO DAY ____________

DATE ____________

TIME ____________



TELEWAR #3

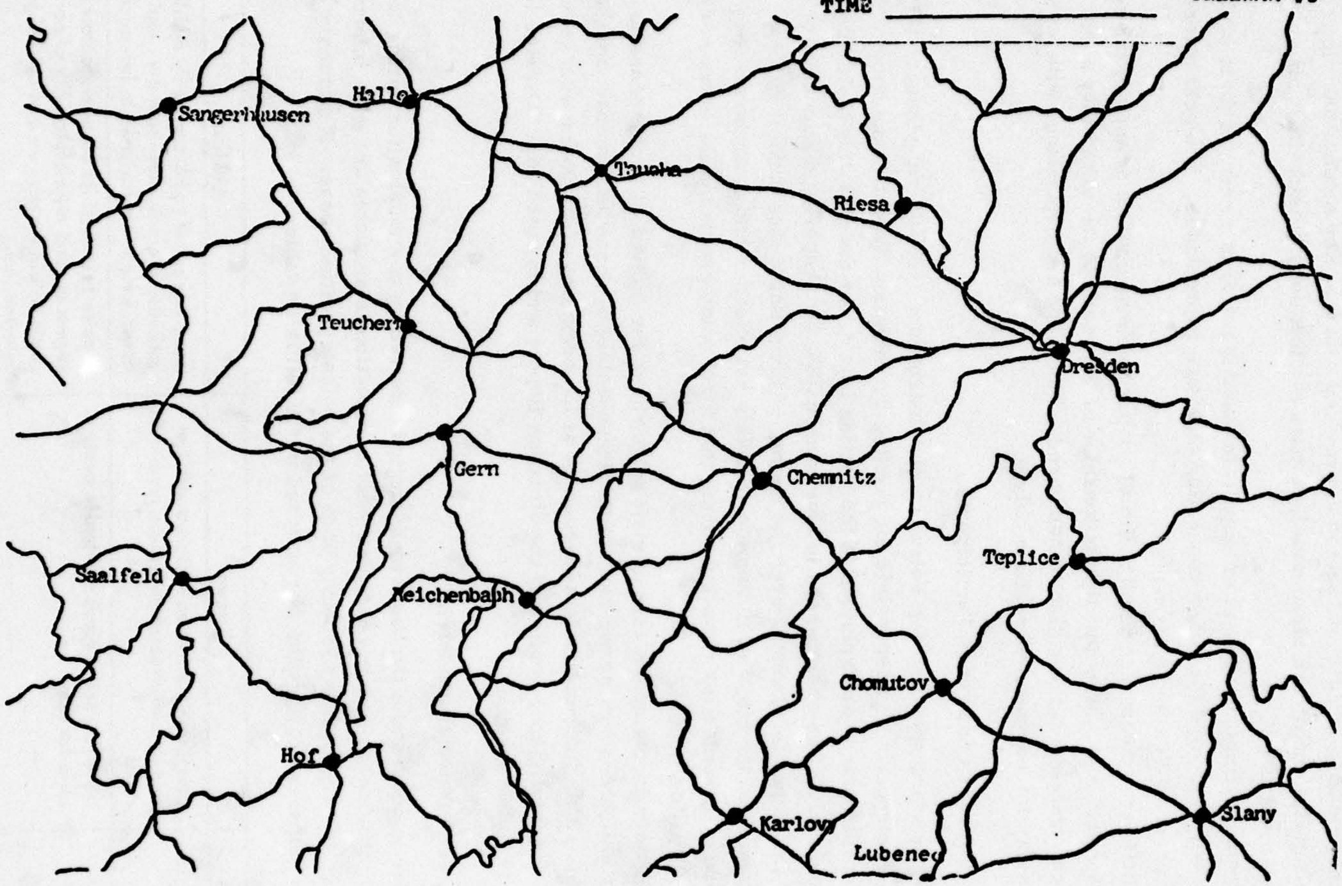


Exhibit B.6.3

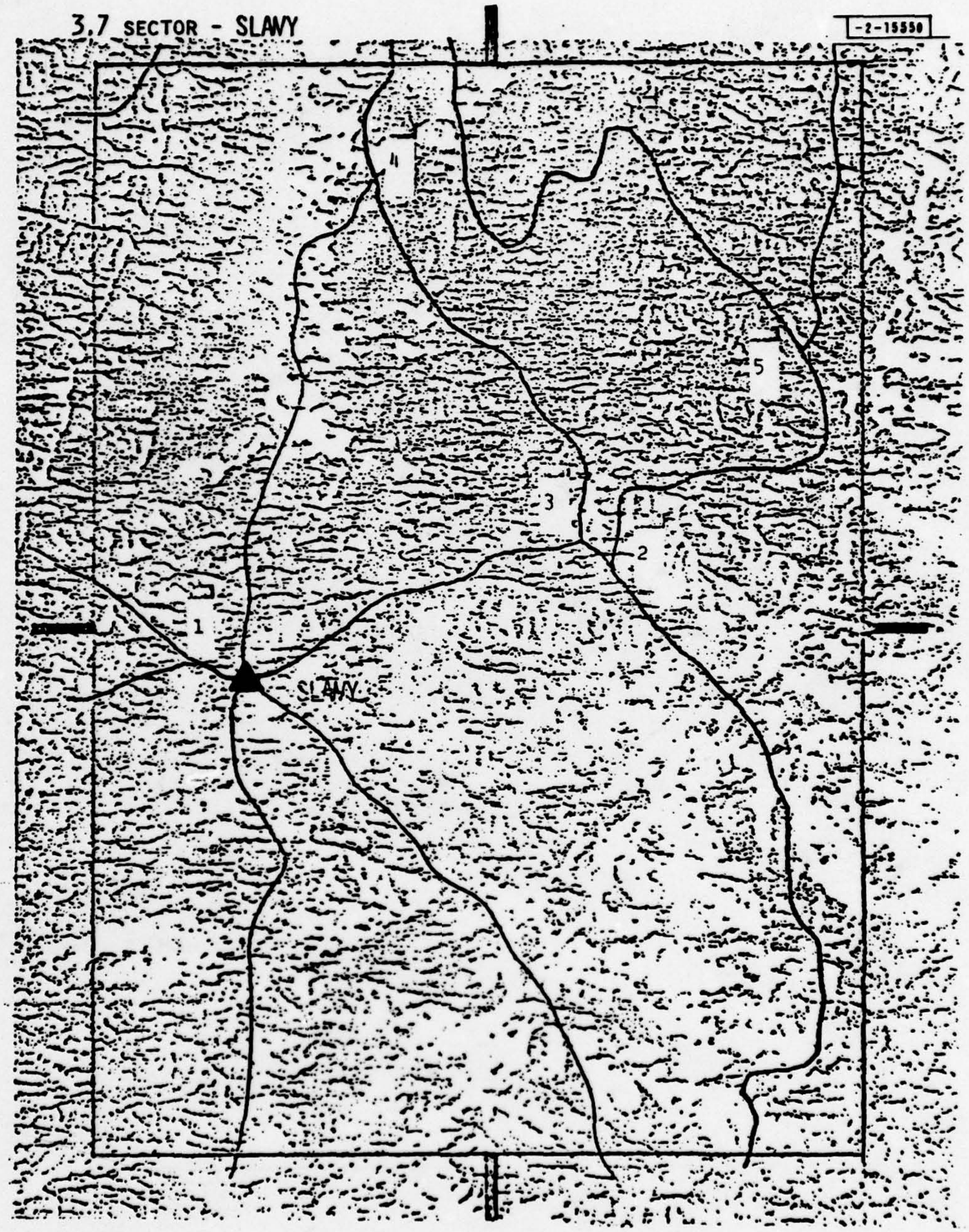


Exhibit B. 6.4a

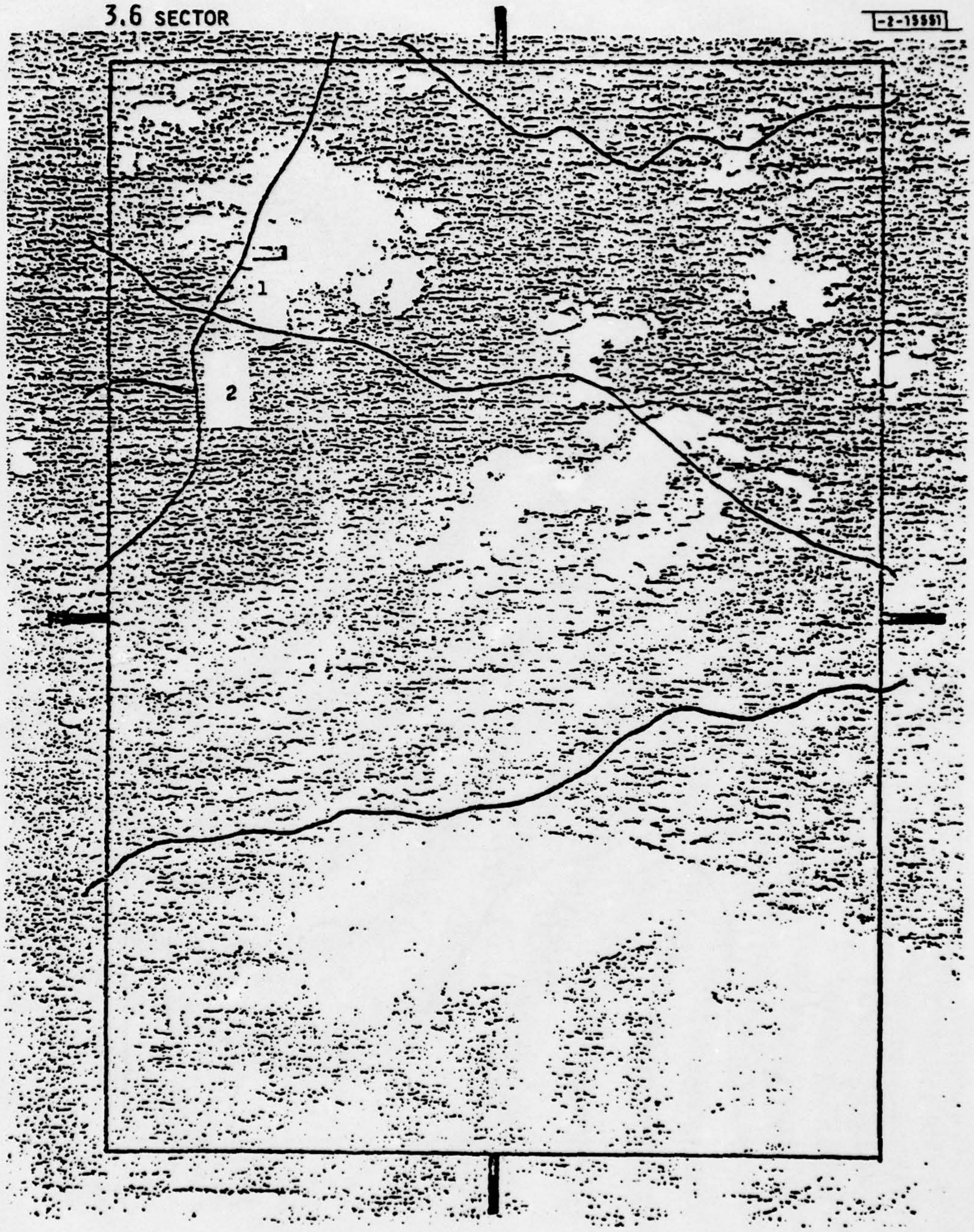


Exhibit B.6.4b

TELEWAR #3

-2-15552

CHAIRPERSON ______________

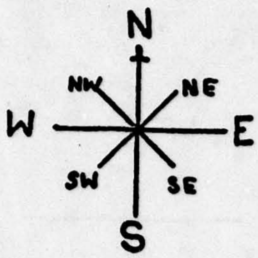
SCENARIO DAY ______________

DATE ______________

FINAL ASSIGNMENT SHEET

LOCATION INITIAL	LOCATION FINAL	ROUTE STRUCTURE (include sector no.'s and towns traversed)	SUM OF INTERSECTIONS

Exhibit B.6.5



TELEWAR 2

DAMAGE REPORT

NAME ____________________

CALL SIGN ________________

SCENARIO DAY ______________

DATE ____________________

TIME ____________________

CITIES

INTERSECTIONS

ROADS

SPECIAL CONDITIONS

-2-19993

Exhibit B.6.6

The procedures used in Telewar are aimed at an orderly transfer of information from staff to planner under direction of the chairperson. A brief example suffices to illustrate the general flow of these procedures:

Assume that one of the objectives for the day is a transfer of units at SLANY to GERN and that a planner decides to begin his routing with the road going NW from the former (Exhibit B.6.3). He indicates this plan to the chairperson who determines from his map (Exhibit B.6.2) that the intended route moves through sectors 3.7 and 3.6. The chairperson communicates this information to participants playing the role of staff who check the detailed maps [Exhibits B.6.4(a) and (b)] and the "Damage Reports" related to those sectors (Exhibit B.6.6). They report to the chairperson and planner whether or not the road is usable and, if so, the value assigned to any intersections traversed by the selected road (two, in this case, with a total value of 2 across sectors 3.7 and 3.6). If the road can be employed, the planner communicates his next intention to the chairperson who alerts appropriate staff personnel. If a road or intersection is impassable at any point, planners, chairperson, and staff are compelled to retrace an intended path in order to find alternative roadways.

After an intact path has been found, the chairperson is required to complete the Final Assignment Sheet (Exhibit B.6.5) by reiterating the sequence of sectors and towns taken by the route and by adding together, with the help of planners and staff, the separate sector-intersection values. As a final step, the chairperson is required to draw on his map (Exhibit B.6.2) an approximation of the selected path. Since the chairperson's actual road information is incomplete, this approximation need only convey the general direction of path within a given sector, but it must traverse sectors and cities in correct sequence.

B.7 "WORD-MATCH"

"Word-Match" is a problem-solving task in which each conferee is provided with a list of words and is required to locate and identify another conferee whose list contains similar words. In actual play, the matching is performed on a word-by-word basis; thus, Conferee No. 1 may say,

> "This is Conferee 1. I have the word 'kid.' Does
> anyone else have 'kid?'"

and hear the reply,

> "This is Conferee 3. I have 'kid.'"

When a match has been found, the conferees write each other's number (No. 3 and No. 1, in this example) next to the word ("kid") on their lists. Play continues until all words on the various lists have been matched or until a designated time has elapsed.

In experiments conducted during the current series, the difficulty of "Word-Match" has been increased slightly by adding words which cannot be matched to the lists.

This task, like "Number-Pass," "Word-Go-Round," and "Path," is easy to learn and requires a minimum of materials. It yields estimation of the ease with which conferees can exchange information. The task may also provide an estimate of intelligibility through careful selection of the words to be matched.

When all words are to be tried, the primary output measure is the total time taken. Since the task was devised to induce collisions and since they occupy a short time relative to the task,

total time taken is not likely to be a sensitive measure of collision-handling procedures. Therefore, greatest importance was attached to participants' ratings.

The people participating in this task had previous experience in teleconferencing and the task is very simple; therefore, the rules of the task were presented in the briefing sessions. The rules are:

(1) Work from the top of your list.
(2) Try only one word at a time.
(3) Do not spell or use the word in a sentence unless it is misunderstood.
(4) Matches must be correctly recorded and reported to count.

A computer program was written and used to generate materials for this task. The list of words used, the number of words per participant, and the mix of non-match, pair-match, and multiple-match words are arbitrary. The version used in Phase III generated eight 10-word lists with 40 non-matches, one octuple-match (all eight participants had the same word somewhere on their list), four quadruple-matches, and eight pair-matches. The section of the master list used, the matching words, the conferees matched, and the position of the words on each list were randomly chosen. The task was stopped after 5 min., ratings were taken with the on-line touch-tone response system, and then matches were reported using the same system. We found that when the octuple-match occurred early in the task, participants' ratings for the system appeared worse than expected.

Therefore, for Phase IV, we placed constraints on the generating program to prevent any occurrence of the octuple-match word in the first or second position of any list. The length of each list was reduced to 5 words and the task was allowed to continue to completion (about 5 min.). To enhance the comparison of systems nearly equal in quality, the generating program was further constrained so that, for any set of runs on a given day, the fine structure of the task (i.e., the position of matching words on a list) remained stable; the words used changed from run to run, as did the list used by each participant. The participants were not told of this stability and none reported noticing it. At no time during Phases III and IV were participants informed as to the structure of the task. Two participants reported knowing the number of non-matches in a list (1) after Phase IV.

A protocol for an eight-person word-match, including the experimenter's script, is attached as Exhibit B.7.1. It is an example of the 5-word lists used in Phase IV.

EXHIBIT B.7.1

A Protocol for an Eight-Person Word-Match

WORD-MATCH ROUND 37 CONFEREE

37	1	2	3	4	5	6	7	8
1	10	23	5	20	9	23	18	7
2	18	14	17	12	17	6	17	13
3	11	18	15	9	21	14	13	12
4	21	19	18	17	16	20	15	20
5	17	17	10	22	20	17	8	17

1	kid	pin	bun	pick	kin	pin	must	kit
2	must	just	gust	king	gust	fun	gust	bust
3	kill	must	rust	kin	pip	just	bust	king
4	pip	pill	must	gust	dust	pick	rust	pick
5	gust	gust	kid	pit	pick	gust	kick	gust

WORD-MATCH ROUND 37 CONFEREE 1

	WORD	MATCH WITH CONFEREE #
1	kid	.....
2	must	.....
3	kill	.....
4	pip	.....
5	gust	.....

EXHIBIT B.7.1 (Continued)

WORD-MATCH ROUND 37 CONFEREE 2

	WORD	MATCH WITH CONFEREE #
1	pin	.....
2	just	.....
3	must	.....
4	pill	.....
5	gust	.....

WORD-MATCH ROUND 37 CONFEREE 3

	WORD	MATCH WITH CONFEREE #
1	bun	.....
2	gust	.....
3	rust	.....
4	must	.....
5	kid	.....

WORD-MATCH ROUND 37 CONFEREE 4

	WORD	MATCH WITH CONFEREE #
1	pick	.....
2	king	.....
3	kin	.....
4	gust	.....
5	pit	.....

WORD-MATCH ROUND 37 CONFEREE 5

	WORD	MATCH WITH CONFEREE #
1	kin	.....
2	gust	.....
3	pip	.....
4	dust	.....
5	pick	.....

EXHIBIT B.7.1 (Continued)

WORD-MATCH ROUND 37 CONFEREE 6

	WORD	MATCH WITH CONFEREE #
1	pin	.....
2	fun	.....
3	just	.....
4	pick	.....
5	gust	.....

WORD-MATCH ROUND 37 CONFEREE 7

	WORD	MATCH WITH CONFEREE #
1	must	.....
2	gust	.....
3	bust	.....
4	rust	.....
5	kick	.....

WORD-MATCH ROUND 37 CONFEREE 8

	WORD	MATCH WITH CONFEREE #
1	kit	.....
2	bust	.....
3	king	.....
4	pick	.....
5	gust	.....

APPENDIX C

TELECONFERENCING QUESTIONNAIRES

APPENDIX C
TELECONFERENCING QUESTIONNAIRES

Several instruments were devised during the project and used to elicit information from participants on various aspects of teleconferencing. Two were designed to retrospectively gather comparative information about systems in Phases I and II. They are called Teleconferencing Questionnaires #1 and #2 and are included as Exhibits C.1 and C.2. A third instrument (Exhibit C.3) was intended to elicit opinion about the characteristics of speech transmitted over systems. The list of characteristics was selected from previous work at BBN on digitally processed speech (BBN Report 3794). The paper version was used in Phase II until the touch-tone response system was implemented. Three descriptors were dropped at the time of conversion, and the remaining 11 incorporated in the participant's script for on-line responding (Exhibit C.5). The list of voice characteristics was not used in Phases III and IV since discrimination among systems on the basis of effects on speech was expected to be very poor.

A set of sentences representing items of interest and including rating response sections was devised and a paper version was used in Phase II. The instructions and sentences are shown in Exhibit C.4. The sentences were refined and incorporated in the participant's script when the touch-tone response system was implemented.

The participant's script (Exhibit C.5) was a multipurpose instrument intended to guide participants through the involved sequence of events in a Phase II session, collect certain comments, store certain responses, and back up the on-line system.

The set of rating items was refined and made more appropriate to Phases III and IV, and, since the script was no longer needed, the result (Exhibit C.6) was shorter and more direct.

EXHIBIT C.1

Introduction to
Teleconferencing Questionnaire #1

A Review of Experimental Conditions to Date

Up to this point, you have participated in at least 10 experimental teleconferencing sessions. During the early sessions, you solved "car pool" problems in four-person groups, somewhat later, in eight person groups, and, very recently, in twelve-person groups. In addition to gaining experience with conferences of different sizes, you have gained experience with two basically different types of telephone systems. One of these, the "analog bridge," is very similar to the common telephone system. The system permits any number of simultaneous speakers to be heard by each other and by all listeners. The second, or "voice control" system, is considerably different from the analog bridge in a number of respects. From the listener's point of view, one of the most prominent of these is that only one speaker can be heard, though several might be attempting to talk. When one speaker has finished, a second may then be heard, though the listener may be aware that early portions of the second speaker's message have been lost.

The Purpose of This Questionnaire

Your perceptions of the ease or difficulty with which conferences can be conducted and problems solved within groups of different sizes using different systems are critical to successful evaluation of various teleconferencing arrangements. Your preferences, if any, among the alternatives are also important.

Exhibit C.1 (Continued)

From time to time, we will ask you to fill out a short questionnaire regarding your perceptions, preferences, and comments. The data provided by you will be used in conjunction with other measures of conference performance with which you are familiar (e.g., solution time, solution quality, tape recordings of the discussions, computer data on the functioning of the phone systems, etc.) in our report of the experimental trials.

Please read and answer all of the questions carefully. When you are finished, put your name in the appropriate space and do either of the following:

- Return the form to Chris.

or

- Keep the form handy and bring it to the next experimental session.

Thank you very much for your continuing cooperation.

BBN/Lincoln

Teleconferencing
Questionnaire #1

PLEASE READ ITEMS CAREFULLY AND COMPLETELY BEFORE ANSWERING

1. Immediately below is a set of five conferencing conditions, each member of which is described by the telephone system employed, the number of conferees and the number of commuters involved in the carpool problem to be solved. You have already served as a subject in each of these conditions.

Index No.	No. of Conferees	Telephone System	No. of Commuters
1	12	Analog Bridge	12
2	12	Voice Control	12
3	8	Analog Bridge	8
4	8	Voice Control	8
5	4	Analog Bridge	8

We ask you to imagine that you will shortly be required to be a subject during a repetition of this set of experimental conditions. This time, however, you are considerably more experienced and have a better perspective on the conferencing situations. As a result, you are able to make an estimate of the rank order of difficulty of the conditions and, further, to make a judgment about how much more difficult or easy one conditon will be than another. In addition, you recognize that, as a result of your accumulating experience, your current perception

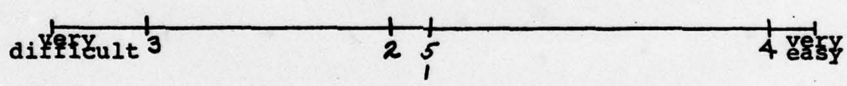
Exhibit C.1 (Continued)

of the relative difficulty of different conditons may not be the same as it was when you were less sophisticated.

It is this current perception, this feeling that you now have about how the conditions would be distributed with respect to difficulty if you were to encounter them again, that we want you to indicate below. Note that you are not being asked to attempt to remember how difficult the conditons seemed at the time, but rather how they now seem in advance of a repetition.

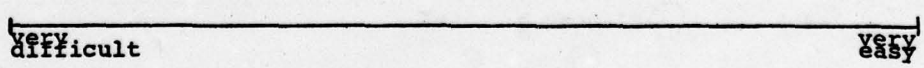
Below, there is a line on which we want you to make your judgments of the relative difficulties of the conference conditions. One end of the line is labelled "very difficult", the other, "very easy". Indicate you judgment of the difficulty of each condition by marking the line at the appropriate point and identifying the mark with the index number associated with that conditon in the above Table. Indicate conditions of equal difficulty by listing associated index numbers in a column below the mark

EXAMPLE:



In this example, a fictitious subject has indicated the belief that condition 3 is quite difficult, that 2 is considerably less difficult than 3 but slightly more difficult than 5 and 1, which are equal. In this subject's view, condition 4 is very much easier than any of the other conditions.

Now it is your turn.



2. Describe, as best you can, the reason(s) why you distributed the conditions as you did.

3. What percentage of the remaining subjects do you believe will distribute the conditions in the same rank order you have? (NOTE: This question concerns order alone, not the distances between marks). Check one.

◯ 0-20%	◯ 41-60%	◯ 81-100%
◯ 21-40%	◯ 61-80%	

Exhibit C.1 (Continued)

4a. How frequently do you believe you can identify conference participants _on the basis of the sounds of their voices_ when using the analog bridge system?

- ◯ almost always
- ◯ frequently
- ◯ infrequently
- ◯ almost never

4b. How frequently do you believe you can identify conference participants _on the basis of the sounds of their voices_ when using the voice control system?

- ◯ almost always
- ◯ frequently
- ◯ infrequently
- ◯ almost never

4c. How important is it to you that you know who is speaking at a given time?

- ◯ very important
- ◯ important
- ◯ not very important
- ◯ very unimportant

5a. Assume that, at some future time, an effort was to be made to determine if a conference involving the solution of car pool could be conducted more efficiently by employing a chairman. The primary task of this chairman would be to eliminate interruptions of one speaker by others. Assume that you were the person chosen. On which of the two systems would you prefer to carry out that role?

- ◯ Analog Bridge
- ◯ Voice Control
- ◯ No Preference

5b. Please explain the reason(s) for the alternative selected above.

PLEASE SIGN YOUR NAME ______________________________

EXHIBIT C.2

TELECONFERENCING
Questionnaire #2

Please read items carefully and completely before answering.

1. Below is the set of five conferencing conditions you rated for relative difficulty three weeks ago, and a copy of your rating form:

CONFERENCE CONDITIONS

Index No.	No. of Conferees	Telephone System	No. of Commuters
1	12	Analog Bridge	12
2	12	Voice Control	12
3	8	Analog Bridge	8
4	8	Voice Control	8
5	4	Analog Bridge	6

YOUR EARLIER RATING

Exhibit C.2 (Continued)

Today you have had experience with a second set of analog bridge and voice control conditions. In this latter set, a delay typical of that which would be experienced during communications involving a satellite was introduced. We would now like you to merge your impressions of these systems with those portrayed in your earlier rating.

Since you have only been in eight-person conferences using the delay conditions, we will eliminate the 12-person and 4-person conditions, leaving the following set for you to judge.

Index No.	No. of Conferees	Telephone System	No. of Commuters
3	8	Analog Bridge	8
4	8	Voice Control	8
6	8	Analog Bridge with Delay	8
7	8	Voice Control with Delay	8

As before, base your judgment on your impression of how the conditions would rank if we were to repeat the experiments in the future.

<u>NOTE</u>: There is no need to maintain either your earlier rank order of Index No.s 3 and 4 or your original spacing. If, in your judgment, either order or spacing has changed in the light of Index No.s 6 and 7, make the change(s) in the rating.

<u>YOUR NEW RATING</u>

very difficult	very easy

2. Distribute Index No.s 3, 4, 6 and 7 on the line below in accord with the relative frequency with which you, as a speaker, feel you would be heard and understood by the rest of the conferees in a future repetition of the experiments.

____

always | never

3. Distribute Index No.s 3, 4, 6, and 7 on the line below in accord with your judgment of the relative ease of interrupting a given speaker when you, as a listener, have something to say.

____

very difficult to interrupt | very easy to interrupt

4. Distribute Index No.s 3, 4, 6 and 7 on the line below in accord with your judgments of the relative frequencies with which you can identify speakers on the basis of the sounds of their voices.

____

always | never

EXHIBIT C.3

Voice Characteristic Checkoff Form

Name____________________

Date__________Time__________

Put a mark on the line if the system made any or all voices sound:

___ as good as on my office telephone.

___ clicky

___ cutoff

___ distorted

___ fuzzy

___ garbled

___ monotonic

___ muffled

___ nasal

___ normal

___ produced by machine

___ squeaky

___ unintelligible

___ unreal

___ _ _ _ _ _ _ _ _ _

___ _ _ _ _ _ _ _ _ _

Use the lines and space at the bottom to indicate any qualities of speech you heard not on the list.

EXHIBIT C.4

Paper Version of Rating Sentences
and Instructions

INSTRUCTIONS FOR SENTENCES

Place a mark under each sentence in the place which expresses your opinion.

You can mark off the line at either end to express an extreme opinion (but make sure we can find it).

The center position is intended to represent neutrality, although several descriptive words are used. If you mark on the centerline, it means you feel equally about both ends of the scale.

Treat each sentence independently. Do not try to make answers match.

Work quickly - read the whole sentence and mark the line.

Do the sentences in order. Do not skip any.

Exhibit C.4 (Continued)

Name________________________ Date__________________ Time______________

This place is	fine	adequate	terrible	to work in.
Speech was	easy	normal	difficult	to understand.
We each get	little	enough	much opportunity	to participate.
This problem was	easier	same	harder	than others.
The system produced	few	average	many	spurious noises
There were	no	usual	many	repeat requests
The handset, buttons, etc. are	easy	normal	hard	to manage.
People talked	rarely	usual	often	at once.
Quality degradation was	less	usual	more	than on other systems.

I had to speak	softer	same	louder	than usual.
This system requires	little	usual	much	fuss to use.
This system changed	no	some	all	voices.
Communication on this system was	easier	same	harder	than on other systems.
I experienced	no	average	great	difficulty being understood.
For this problem, I	like	don't mind	dislike	this system.
This system is	better	same	worse	than most others
I missed	few	some	many	words.
This system is	excellent	average	terrible	in overall quality

EXHIBIT C.5

Participant's Script

LAST NAME____________________ ROOM #____ EXT____ DIAL_____ RUN #______

- - - *Dial-up*

- - - *Read problem*

```
                  '   hard   '  average  '   easy   '
This problem    ---+---+---+---+---+---+---+---+---+---   to solve.
will be
```

- - - *Do word-go-round. Check how voices sound.*

____fuzzy	____clicky	____cutoff	____muffled
____nasal	____garbled	____monotonic	____squeaky
____unintelligible	____unreal	____produced by machine	

```
                  '   easy   '  average  '   hard   '
Working on this ---+---+---+---+---+---+---+---+---+---
system will be
```

- - - *Do problem. Check how voices sound.*

____fuzzy	____clicky	____cutoff	____muffled
____nasal	____garbled	____monotonic	____squeaky
____unintelligible	____unreal	____produced by machine	

What was easiest about that problem?

____

____

What was hardest about that problem?

____

____

- - - *Turn Page*

```
                  '   bad    '  average  '   good   '
Overall, this   ---+---+---+---+---+---+---+---+---+---
system was
```

This room is	quiet	normal	noisy	to work in.
Speech was	easy	normal	difficult	to understand.
I had	insufficient	sufficient	ample	time to speak.
The system produced	few	some	many	spurious sounds.
There were	many	some	few	repeat requests.
The handset, buttons, etc.were	hard	normal	easy	to manipulate.
People talked	rarely	sometimes	often	at once.
I had to speak	softer	same	louder	than usual.
My contribution was	great	good	poor	to this problem.
This system requires	little	usual	much	effort to use.
This system changed	many	some	few	voices.
This system was better than	few	some	many	other systems.
My speech was	often	usually	rarely	understood.
This problem was	dull	average	interesting	
Group performance was	poor	good	great	for this problem.
Communication is	better	same	worse	than free-air.
Work in this problem was	helped	unaffected	hindered	by the handset, buttons, etc.
I missed	many	some	few	words.
We had	little	enough	plenty	time for the problem.

EXHIBIT C.6

NAME DATE TIME RUN #

ROOM EXT. DIAL

Overall, this system was	good	average	bad	
	* 1 2 3	4 5 6	7 8 9 #	
This system requires	little	usual	much	listening effort.
	* 1 2 3	4 5 6	7 8 9 #	
The collision signal was	helpful	neutral	unhelpful	in performing the task.
	* 1 2 3	4 5 6	7 8 9 #	
The collision signal was	pleasing	neutral	annoying	during the task.
	* 1 2 3	4 5 6	7 8 9 #	
When I wanted to talk, I had	little	some	much	difficulty gaining the floor.
	* 1 2 3	4 5 6	7 8 9 #	
Speech was	easy	normal	difficult	to understand.
	* 1 2 3	4 5 6	7 8 9 #	
This system changed	few	some	many	voices.
	* 1 2 3	4 5 6	7 8 9 #	
The system produced	few	some	many	spurious sounds.
	* 1 2 3	4 5 6	7 8 9 #	
I missed	few	some	many	words.
	* 1 2 3	4 5 6	7 8 9 #	
There were	few	some	many	repeat requests.
	* 1 2 3	4 5 6	7 8 9 #	
I had to speak	softer	same	louder	than usual.
	* 1 2 3	4 5 6	7 8 9 #	

APPENDIX D

ON-LINE DATA ACQUISITION AND PROCESSING

APPENDIX D
ON-LINE DATA ACQUISITION AND PROCESSING

Much of the rating data in Phase II and all of it in Phases III and IV was acquired by a computer at Lincoln Laboratory, sent electronically to BBN, and processed there with a series of utility and specially written programs. This appendix describes the process in its most advanced form (a Phase IV run) and provides examples of the output at various stages.

D.1 DATA ACQUISITION

Word Match 37 identifies an eight-person, five-word, run-to-completion conference on an Analog-Bridge Circuit in Phase IV. The materials used by the participants are shown in Appendix B as Exhibit B.7.1. When the participants complete the task, the telephone system is switched to response mode. The experimenter reads the initial phrase of each rating item (shown in Exhibit C.6) and the participants respond by pressing buttons on their touch-tone pads. After the last rating item, the "match with conferee #" column is input in the same manner. Then a new circuit is selected and the process repeated. At the end of the session, rating and match data for as many as seven circuits have been acquired and stored.

D.2 DATA TRANSMISSION AND CONDITIONING

The entire set of data for the session is sent as a single message over the ARPANET to an electronic mailbox. It is then converted to an ordinary data file in a user directory on a BBN TOPS-20 computer system. This file is then processed with an interactive editing program, extraneous matter is stripped away, and the various sections of the data isolated and constituted as individual data files. Exhibit D.1 shows the data file for the ratings for WM 37, and Exhibit D.7 shows the "match with" data file.

The final step in the conditioning process is a translation of pound signs (#) and asterisks (*) to numeric quantities (11 and 1, respectively). Various descriptive statistics are then computed and a frequency plot is generated. Exhibit D.2 shows first the original and conditioned data for each conferee, then the frequency plot (items are rows and ratings are columns), and then the statistics. The data conditioning program also formats and outputs a data file (Exhibit D.3) for use by other programs.

D.3 DATA PROCESSING

The file and others like it are combined, normalized, and output (Exhibit D.4), and used as input to other programs. One program performs the Wilcoxon matched-pairs/signed-ranks test; a sample of output is shown in Exhibit D.5. Another program creates a crude graphical comparison of system means for each question; a sample of output is shown in Exhibit D.6. Many of the programs have varieties of forms for special applications; these are not shown.

```
SAW 34554215115
TIE 21553111135
JOY 56553515116
WYN 47554523235
COO 77556656557
CHA 64556437345
BOY 31554114245
COL 66555636475
```

Exhibit D.1.

Data File of ratings for word-match 37. Each row presents the responses for one participant, columns correspond to rating items.

```
M37  8
SAW 3  4  5  5  4  2  1  4  1  1  5
       4  5  6  6  5  3  2  5  2  2  6
TIE 2  1  5  5  3  1  1  1  1  3  5
       3  2  6  6  4  2  2  2  2  4  6
JOY 5  6  5  5  3  5  1  5  1  1  6
       6  7  6  6  4  6  2  6  2  2  7
WYN 4  7  5  5  4  5  2  3  2  3  5
       5  8  6  6  5  6  3  4  3  4  6
COU 7  7  5  5  6  6  5  6  5  5  7
       8  8  6  6  7  7  6  7  6  6  8
CHA 6  4  5  5  6  4  3  7  3  4  5
       7  5  6  6  7  5  4  8  4  5  6
BOY 3  1  5  5  4  1  1  4  2  4  5
       4  2  6  6  5  2  2  5  3  5  6
COL 6  6  5  5  5  6  3  6  4  7  5
       7  7  6  6  6  7  4  7  5  8  6
```

Exhibit D.2 (page 1)

Original and conditioned data for word-match 37. This output facilitates two comparisons: (1) between the original data message and the data as read by the conditioning program, and (2) between the original and the conditioned data.

RM37

	M	1	2	3	4	5	6	7	8	9	10	11	
1				1	2	1	1	2	1				5.50
2			2			2		2	2				5.50
3							8						6.00
4							8						6.00
5					2	3	1	2					5.38
6			2	1		1	2	2					4.75
7			4	1	2		1						3.13
8			1		1	2	1	2	1				5.50
9			3	2	1	1	1						3.38
10			2		2	2	1		1				4.50
11							6	1	1				6.38
FREQ	0	0	14	5	10	12	30	11	6	0	0	0	
MNFREQ	0.0	0.0	1.3	0.5	0.9	1.1	2.7	1.0	0.5	0.0	0.0	0.0	
	M	1	2	3	4	5	6	7	8	9	10	11	

Exhibit D.2 (page 2)

Frequency plot of data for word-match 37. Rows represent rating items and columns represent possible ratings (M = missing data); entries indicate the frequency of each rating for each item. The mean rating for each item appears at the right, and the total frequency and mean frequency for each rating appear under the plot.

RM37

#	MAX	MIN	RNG	X	VAR	SD	AD	MED	Q1	Q3	SMQR
1	8	3	5	5.50	3.14	1.77	1.50	5.50	4.00	7.00	1.50
2	8	2	6	5.50	6.00	2.45	2.00	6.00	3.50	8.00	2.25
3	6	6	0	6.00	0.00	0.00	0.00	6.00	6.00	6.00	0.00
4	6	6	0	6.00	0.00	0.00	0.00	6.00	6.00	6.00	0.00
5	7	4	3	5.38	1.41	1.19	0.97	5.00	4.50	7.00	1.25
6	7	2	5	4.75	4.50	2.12	1.81	5.50	2.50	7.00	2.25
7	6	2	4	3.13	2.13	1.46	1.16	2.50	2.00	4.00	1.00
8	8	2	6	5.50	3.71	1.93	1.50	5.50	4.50	7.00	1.25
9	6	2	4	3.38	2.27	1.51	1.22	3.00	2.00	5.00	1.50
10	8	2	6	4.50	4.00	2.00	1.50	4.50	3.00	6.00	1.50
11	8	6	2	6.38	0.55	0.74	0.56	6.00	6.00	7.00	0.50

Exhibit D.2 (page 3)

Descriptive statistics for word-match 37. Rating items are rows and statistics are columns.

Column Heading	Entry
#	Rating Item
MAX	Maximum Rating
MIN	Minimum Rating
RNG	Range of Ratings
X	Mean of Ratings
VAR	Variance
SD	Standard Deviation
AD	Average Deviation
MED	Median Rating
Q1	First Quartile
QJ	Third Quartile
SMQR	Quartile Deviation

Note that statistics are performed on conditioned data.

```
SM37  8
SAW   4  5  6  6  5  3  2  5  2  2  5
TIE   3  2  6  6  4  2  2  2  2  4  6
JOY   6  7  6  6  4  6  2  6  2  2  7
WYN   5  8  6  6  5  6  3  4  3  4  6
COO   8  8  6  6  7  7  6  7  6  6  3
CHA   7  5  6  6  7  5  4  8  4  5  6
BOY   4  2  6  6  5  2  2  5  3  5  6
COL   7  7  6  6  6  7  4  7  5  8  6
```

Exhibit D.3

Conditioned output data file for word-match 37. Subjects are rows and items are columns in this file, which is used as input to special-purpose computational programs. The first row ("SM37 8") contains the identification number of the run and the number of conferees.

```
CHA   6  6  4  6  7  4  4  5  7  7  6
BOY   4  3  4  6  4  2  2  3  3  3  6
COL   7  8  4  7  8  6  3  6  4  8  6
SM33  8
SAW   6  6  4  2  7  4  2  2  5  6  6
TIE   8  2  8  6  8  2  2  2  8  8  6
JOY   5  6  4  7  6  6  3  6  5  5  6
WYN   9  8  3  4  5  7  3  5  9  7  6
COO   7  7  9  7  7  7  6  8  6  7  6
CHA   5  6  4  6  7  4  4  5  8  8  7
BOY   3  2  4  6  2  2  2  2  3  2  6
COL   4  5  4  4  6  6  3  3  6  7  6
SM34  8
SAW   2  2  2  2  4  2  1  3  3  3  6
TIE   3  2  6  6  4  2  2  2  4  5  6
JOY   5  7  4  6  4  7  7  7  4  4  6
WYN   4  6  6  6  3  6  3  6  2  2  6
COO   5  6  6  6  5  6  6  7  5  6  6
CHA   5  6  5  6  7  5  4  5  7  7  7
BOY   4  3  6  6  3  2  2  4  2  2  6
COL   3  5  3  3  3  3  3  3  5  6  6
SM35  8
SAW   6  6  3  2  7  3  2  3  5  5  6
TIE   3  2  3  5  6  2  2  2  4  5  6
JOY   4  7  5  5  4  6  4  5  4  5  6
WYN   4  6  3  4  2  6  2  3  3  5  6
COO   5  6  5  6  5  7  7  7  7  7  6
CHA   4  4  4  6  7  4  4  4  6  7  6
BOY   2  2  6  6  2  2  2  2  2  2  6
COL   6  8  7  5  9  6  3  3  7  9  6
SM37  8
SAW   4  5  6  6  5  3  2  5  2  2  6
TIE   3  2  6  6  4  2  2  2  2  4  6
JOY   6  7  6  6  4  6  2  6  2  2  7
WYN   5  8  6  6  5  6  3  4  3  4  6
COO   8  8  6  6  7  7  6  7  6  6  8
CHA   7  5  6  6  7  5  4  8  4  5  6
BOY   4  2  6  6  5  2  2  5  3  5  6
COL   7  7  6  6  6  7  4  7  5  8  6
.
```

Exhibit D.4 (page 1)

Section of data file containing "SM37" and other files.

```
SM39  8
SAW  0.78 2.94 2.39-1.11 1.17 0.11-0.11 1.17 2.06 2.56 0.06
TIE -0.94 0.61-2.39-0.56 0.00 0.00 0.06 0.00 0.00-1.39 0.00
JOY  0.06-0.78-0.61 0.56 0.22-0.06-0.89-0.94-1.39-1.50-0.06
WYN -1.67-1.50 0.72 0.44-1.50-2.50-0.61-0.67-1.33 1.22-0.33
COO -0.11-0.39-1.11 0.28 0.94-0.33-0.11 0.17-0.72-0.33-0.33
CHA  0.61 1.28 1.22 0.00-0.83 0.33-0.06 2.61 0.39 0.06 0.61
BOY -0.78-0.17 0.78 0.00-1.11 0.00 0.00-0.56-0.39-0.39 0.00
COL -1.33-1.78 0.83 0.50-1.89-2.06-0.83-1.78-1.33-2.22 0.00
SM42  8
SAW -2.22-3.06-1.61-0.11-0.83-0.89-0.11-0.83-0.94-0.44 0.06
TIE  0.06 0.61-2.39-0.56 0.00 0.00 0.06 0.00 0.00-0.39 0.00
JOY -0.94-0.78-0.61 0.56-1.78-0.06-0.89-1.94-1.39-1.50-0.06
WYN -1.67-0.50 0.72 0.44 0.50 0.50 0.39 0.33-1.33-0.78-0.33
COO -1.11-1.39-2.11 1.28-0.06-1.33-1.11-2.83-0.72-0.33-0.33
CHA -0.39 0.28-0.78 0.00-0.83 0.33-0.06 0.61-1.61-0.94-0.39
BOY -0.78-0.17-1.22 0.00 0.89 0.00 0.00-0.56 0.61-0.39 0.00
COL -1.33-1.78 0.83 0.50-1.89-1.06 0.17 0.22-2.33-2.22 0.00
CM29  8
SAW -0.72-1.06 1.39 2.89-0.83 0.11 0.39 1.17-1.44-1.44-0.44
TIE -1.44-0.39 1.61 0.44-1.50 0.00-0.44 0.00-2.00-0.89 0.00
JOY  0.56 0.72 1.39-0.44 1.22-0.06 0.11 0.56 0.11 0.50 0.44
WYN  0.33 0.50 0.72 0.44 1.50-1.50-0.11 0.83-1.33-1.28 0.17
COO  0.39 0.11-0.11-0.72-0.56 0.67-0.11 0.17 0.28-0.33 0.67
CHA  0.61-1.22 1.22 0.00 0.17 0.33-0.06 1.61-1.11-0.94-0.39
BOY  1.22-0.17 0.78 0.00 1.39 0.00 0.00 1.94 0.11 1.61 0.00
COL  1.67 1.72 0.83 0.50 1.11 0.44 0.17 2.22 1.17 2.28 0.00
CM30  8
SAW  0.78 0.44-1.61-1.11-1.33-0.89-0.11 2.17 0.06 0.56 0.06
TIE -1.44 0.11-1.89-0.56-1.00 0.00 0.06 0.00 0.00-0.89 0.00
JOY -0.44-0.28-0.61 0.06-0.78-0.06 0.11-0.94 0.61 0.50-0.06
WYN  1.33 0.00-1.28-0.56 1.50-0.50-0.11-1.17 0.17-1.28 0.67
COO -0.11 0.11-0.61-0.22-0.06-0.33-0.11-0.33 0.28 0.17 0.17
CHA -0.39-0.22-0.78 0.00-0.33 0.83 0.44 0.11-0.61-0.94-0.39
BOY -0.78-0.17 0.78 0.00-0.11 0.00 0.00-0.56-0.39-0.39 0.00
COL -0.83-0.28-1.67-1.00-0.89-1.06-0.33-0.78-0.83-1.72 0.00
CM31  8
SAW -2.22-1.06-0.61 0.89-0.33 0.11-0.11-0.33-0.44-0.44 0.06
TIE -2.44-0.39 0.61 0.44-1.50 0.00 0.06 0.00-2.00-0.89 0.00
JOY  1.56 0.22 1.39-0.44 0.22-0.06-0.39 2.56-0.39-0.50-0.06
WYN -1.17-1.00 0.72 0.44-1.50-1.00-0.11-0.67-1.83-0.28 0.17
COO -0.11 0.11-0.61-0.22-1.06 0.17 0.89 0.67-0.22-0.33 0.17
CHA  0.11 0.28 0.22 0.00-0.83-0.17-0.06 0.61-1.11-0.94 0.11
BOY -0.28-0.17 0.78 0.00-0.61 0.00 0.00-0.56-0.39-0.39 0.00
COL -0.83-1.28 0.83 0.50 0.11-0.56 0.17 1.22 0.67-1.22 0.00
```

Exhibit D.4 (page 2)

Section of data file containing "SM37," combined with "SM29," named "CM29," and normalized with other files.

1	SM39	SM32	8	7	0.0000	0.0000	.02
1	SM42	CM29	8	8	3.5000	0.0000	.05
1	SM42	SM32	8	8	0.0000	0.0000	.01
1	SM42	SM33	8	7	0.0000	0.0000	.02
1	CM29	CM31	8	8	3.5000	0.0000	.05
1	CM29	SM32	8	6	0.0000	0.0000	.05
1	CM29	CM34	8	8	2.0000	0.0000	.02
1	CM30	SM32	8	8	0.0000	0.0000	.01
1	CM30	SM33	8	7	0.0000	0.0000	.02
1	CM31	SM32	8	8	1.5000	0.0000	.02
1	SM32	CM34	8	8	0.0000	0.0000	.01
1	SM32	CM35	8	8	0.0000	0.0000	.01
1	SM33	CM35	8	8	2.5000	0.0000	.05
2	SM42	SM32	8	7	2.0000	0.0000	.05
2	CM29	SM32	8	7	1.0000	0.0000	.05
2	CM30	SM32	8	8	2.0000	0.0000	.02
2	SM32	CM35	8	7	1.0000	0.0000	.05
3	SM42	CM29	8	6	0.0000	0.0000	.05
3	SM42	CM31	8	6	0.0000	0.0000	.05
3	CM29	CM30	8	7	0.0000	0.0000	.02
3	CM29	CM35	8	8	1.0000	0.0000	.02
3	CM30	CM31	8	6	0.0000	0.0000	.05
3	CM30	CM34	8	7	0.0000	0.0000	.02
3	CM31	CM35	8	8	2.5000	0.0000	.05
5	SM39	SM32	8	8	0.0000	0.0000	.01
5	SM39	SM33	8	7	2.0000	0.0000	.05
5	SM42	SM32	8	8	0.0000	0.0000	.01
5	SM42	SM33	8	8	3.0000	0.0000	.05
5	CM29	CM31	8	7	1.5000	0.0000	.05
5	CM29	SM32	8	8	4.0000	0.0000	.05
5	CM29	CM34	8	7	2.0000	0.0000	.05
5	CM30	SM32	8	8	0.0000	0.0000	.01
5	CM30	SM33	8	8	4.0000	0.0000	.05
5	CM31	SM32	8	7	0.0000	0.0000	.02
5	CM31	SM33	8	8	1.0000	0.0000	.02
5	SM32	CM34	8	8	0.0000	0.0000	.01
5	SM32	CM35	8	8	2.0000	0.0000	.02
5	SM33	CM34	8	8	1.0000	0.0000	.02
5	SM33	CM35	8	8	2.0000	0.0000	.02

Exhibit D.5

Sample of output showing results of Wilcoxon matched-pairs signed-ranks test. Left column is rating item number, next are the names of pairs compared, number of observations, number of non-tied observations, the calculated statistic (T), the standard score (Z) computed when N>25, and the level of significance (p). Non-significant results are not shown.

SM39	-0.42	0.03	0.23	0.01	-0.38	-0.56	-0.32	0.00	-0.34	-0.25	-0.01
SM42	-1.05	-0.85	-0.90	0.26	-0.50	-0.31	-0.19	-0.63	-0.96	-0.87	-0.13
CM29	0.33	0.03	0.98	0.39	0.31	-0.00	-0.01	1.06	-0.53	-0.06	0.06
CM30	-0.23	-0.04	-0.96	-0.42	-0.38	-0.25	-0.01	-0.19	-0.09	-0.50	0.06
CM31	-0.67	-0.41	0.42	0.20	-0.69	-0.19	0.06	0.44	-0.71	-0.62	0.06
SM32	2.20	1.40	0.48	0.39	1.75	1.06	0.31	0.12	1.54	1.13	0.12
SM33	1.20	0.40	-0.02	-0.36	1.13	0.44	0.06	-0.25	1.91	1.63	-0.01
CM34	-0.80	-0.35	0.17	-0.05	-1.00	-0.13	0.06	-0.06	-0.71	-0.56	-0.01
CM35	-0.55	-0.22	-0.40	-0.42	-0.25	-0.06	0.06	-0.50	-0.09	0.13	-0.13

Exhibit D.6 (page 1)

Means of normalized data for runs (rows) for rating items (columns).

MEANS OF RUNS

```
1 I                     HI                             I
1 I                   B EAD  C    G     F              I
   1----2----3----4----5----6----7----8----9----10---11

2 I                       I C                          I
2 I                    B EHDA G    F                   I
   1----2----3----4----5----6----7----8----9----10---11

3 I                    D      F                        I
3 I                    B  IGHAE C                      I
   1----2----3----4----5----6----7----8----9----10---11

4 I                          F                         I
4 I                          E                         I
4 I                      I   C                         I
4 I                      DGHAB                         I
   1----2----3----4----5----6----7----8----9----10---11

5 I                       I                            I
5 I                       D                            I
5 I                    HEBA  C   G  F                  I
   1----2----3----4----5----6----7----8----9----10---11

6 I                        I                           I
6 I                       DH                           I
6 I                      ABEC G  F                     I
   1----2----3----4----5----6----7----8----9----10---11

7 I                         I                          I
7 I                        DH                          I
7 I                        CG                          I
7 I                       ABEF                         I
   1----2----3----4----5----6----7----8----9----10---11

8 I                        HF                          I
8 I                     BIGDA E  C                     I
   1----2----3----4----5----6----7----8----9----10---11

9 I                     H  I                           I
9 I                    BECAD       F G                 I
   1----2----3----4----5----6----7----8----9----10---11
```

MEANS OF RUNS (continued)

```
10 I                         H                           I
10 I                       BEDACI    F  G                I
     1----2----3----4----5----6----7----8----9----10---11

11 I                          I                          I
11 I                          HF                         I
11 I                          GE                         I
11 I                          BD                         I
11 I                          AC                         I
     1----2----3----4----5----6----7----8----9----10---11
```

1

```
A   SM39
B   SM42
C   CM29
D   CM30
E   CM31
F   SM32
G   SM33
H   CM34
I   CM35
```

Exhibit D.6 (pages 2,3)

Sample of lineprinter plotting. This plot shows the means of normalized data for runs (from Exhibit D.4, page 1) plotted for each rating item (1-11, identified in the left margin) with the original data scale (1-11) on the abscissa. The entries stack vertically to prevent overprinting. The mean for each rating item equals "6" on this plot. Favorable responses plot less than 6, i.e., to the left.

D.4 WORD-MATCH SCORING

The "match with" data file (Exhibit D.7) from the participant's responses and the map of the participant word matrix (Exhibit D.8) are input to a scoring program which produces the page shown in Exhibit D.9. This page indicates any errors made in the word-match task and is the key to identifying the cause or type of errors. Note, for example, that one participant (TIE) is scored with one error: a match-wrong of the fourth word (19 on corresponding position above) with participant #7 (BOY) who does not have that number word. Inspection of TIE's data sheet shows the correct response written on the fourth line; the error was in responding and not in communicating on the system.

```
SAW 33953
TIE 66777
JOY 94711
WYN 58539
COO 46194
CHA 29395
BOY 22839
COL 97461
```

Exhibit D.7

A sample of a "match-with" data file, showing for each conferee (rows), for each item on his/her word list (columns), the conferee number matched-with, or the numeral "9," indicating "no-match."

1	10	23	5	20	9	23	18	7
2	18	14	17	12	17	6	17	13
3	11	18	15	9	21	14	13	12
4	21	19	18	17	16	20	15	20
5	17	17	10	22	20	17	8	17

Exhibit D.8

The participant word matrix shows, for each person (columns), for each list position (rows), the number of the word used. This file is abstracted by an editing process from the output of the stimulus-generating program.

1 DWM37

1	10	23	5	20	9	23	18	7
2	18	14	17	12	17	6	17	13
3	11	18	15	9	21	14	13	12
4	21	19	18	17	16	20	15	20
5	17	17	10	22	20	17	8	17

Participant Word Matrix

RWM37

	SAW	TIE	JOY	WYN	COO	CHA	BOY	COL
1	3	6	9	5	4	2	2	9
2	3	6	4	8	6	9	2	7
3	9	7	7	5	1	2	8	4
4	5	7	1	3	9	8	3	6
5	8	7	1	9	4	5	9	1

Match-With Matrix

SCORES FOR CONFEREES

	MR	MW	NMR	NMW	WNT
SAW	4	0	1	0	0
TIE	4	1	0	0	0
JOY	4	0	1	0	0
WYN	4	0	1	0	0
COO	4	0	1	0	0
CHA	4	0	1	0	0
BOY	4	0	1	0	0
COL	4	0	1	0	0
TOTALS:	32	1	7	0	0

TOTAL RIGHT= 39
TOTAL WRONG= 1
TOTAL TRIED= 40
PERCENT TRIED=100.00
PERCENT CORRECT= 97.50
RANGE= 0
RIGHT PER MINUTE=

MULTIPLES	WORD	SIZE	MATCHES
	17	8	8
	18	4	4
	20	4	4

Exhibit D.9

Sample of word-match scoring program output. "MR" = match right, "MW" = match wrong, "NMR" - no-match right, and "WNT" = was not tried.

D.5 AUDIT TRAILS AND STATISTICAL PACKAGE

Exhibit D.5.1 presents a sample output generated by the data reduction program upon the completion of a conference experiment. The first part of the output is an audit trail showing the distribution of speech generated in the conference as a function of time. This particular conference involved 12 participants and lasted for 770 sec. It used the SI voice-controlled signal selection technique. Time is represented in horizontal bands with tick marks every 10 sec. Within each band are rows for each participant, which are labeled by the columns of numbers on the left and right margins. The row labeled "20" is used to note events marked by the experimenters, such as the actual starting time of the conferencing problem. Each character in a row indicates the "state" of a given participant's telephone line during a 1-sec interval. If the participant was the selected "speaker" during the interval, a "1" appears. If he was the "interrupter," a "-" appears. If a speaker became an interrupter, or vice versa, a "+" appears. If he produced signal energy above the speech activity threshold, but was not selected as either speaker or interrupter, a "o" appears. If he was silent, no mark appears. The numbers below the tick marks at the bottom of each band show the time since the startup of the conference.

Following the audit trail, a series of summary statistics are printed. Each is identified by a title. The statistics are based on samples taken every 20 msec from each participating phone line. Each sample is categorized as belonging to one of eight states, as follows:

State	Meaning
0	No speech detected
1	Speech above threshold detected but channel not selected
2	"Speaker" selected but currently not speaking
3	"Speaker" selected and currently speaking
4	"Interrupter" selected but currently not speaking
5	"Interrupter" selected and currently speaking
6	"Speaker" and "interrupter" selected but neither speaking (not a meaningful category at this time)
7	"Speaker" and "interrupter" selected and both speaking

```
Filename: ts5
number of header words = 4
start data coll at Fri Oct 14 09:18:26 1977
trail off length = 500 ms
size of filled silence gaps = 300 ms or less
size of ignored spurts = 200 ms or less
defn of talking: bit0 = 1

from beginning to end

--------------------------------------------------------------------------------

Each column is 50 * 20ms

audit chars:
        |       speaker
        -       interrupter
        +       both speaker and interrupter
        o       above threshold, not speaker or interrupter
        (space) none of the above

 1:                                                          |||            :1
 2:          ||----               -- --           || o         +|           :2
 3:           o                         o                           ||  +|  :3
 4:                                                ||||              ||||   :4
 5:           oo ||                                     |||                 :5
 6:           --o ||||||||  o|                         +++-+|-+|+||     ||  :6
 7:           ||||--  o -+||| ||       ---    +||                           :7
 8:                   ---o    || ||||         --      --                    :8
 9:                        --- ||      o||      --  - || ||                 :9
10:           oo         ||                 ||| ||      -                   :10
11:          | +| o ---   --        +|||||||                                :11
12:                                                                         :12
20:                                                                         :20
             !         !         !         !         !         !         !
         10.00     20.00     30.00     40.00     50.00     60.00     70.00
  136.86  beep on
  137.12  beep off at addr 0,7126 (oct) in file
 1:                                                                         :1
 2:   ||||||+||                                                             :2
 3:                             |||                                         :3
 4:                                +|                                       :4
 5:                                      ||      ---           ||  |||||||| :5
 6:  |                    ||||||| |||| || |||||||||||||||||-+|||   || --    :6
 7:                                         ||                              :7
 8:            |||                             ||          ||               :8
 9:  ||   +|           |||                       ||            -+| --       :9
10:              --                               +|       |                :10
11:              ||||||                             +|                      :11
12:                                                      ||                 :12
20:                                                                    oo   :20
             !         !         !         !         !         !         !
         80.00     90.00    100.00    110.00    120.00    130.00    140.00

 1:            ||||||||-+||||||                                             :1
 2:                                                                         :2
 3:                                                                         :3
 4:                                                                         :4
 5:  |||||||                                                                :5
 6:                                                                  |      :6
 7:                                                                         :7
 8:                                                                    |||| :8
 9:                                           ||||||| |||||                 :9
10:    --                                                ||||||||+--        :10
11:                   ||      |||||| ||||||                                 :11
12:                                                                         :12
20:                                                                         :20
             !         !         !         !         !         !         !
        150.00    160.00    170.00    180.00    190.00    200.00    210.00
```

```
 1:                                                                        :1
 2:                                                                        :2
 3:                                                  |||||||||             :3
 4:                                                            +|||||||||  :4
 5:                                                                        :5
 6:                         ||||||| ||||||                                 :6
 7:             |   |||||||                                                :7
 8: || ||||||                                                              :8
 9:                                                             |          :9
10:                                                                        :10
11:                                                                        :11
12:                                        ||||||||                        :12
20:                                                                        :20

             !         !         !         !         !         !         !
        220.00    230.00    240.00    250.00    260.00    270.00    280.00

 1:                                        --      ||||||                  :1
 2:         ||||||||| |||||         |||                                    :2
 3:                                                                        :3
 4:  ||||                           --+|   |||||||                -        :4
 5:                                                                        :5
 6:                          ||||||||                                      :6
 7:                                                                        :7
 8:                                     ||                  ||| ||         :8
 9:                                       ||                               :9
10:                                                               |        :10
11:                                                               ||||     :11
12:                                                                   |||| :12
20:                                                                        :20

             !         !         !         !         !         !         !
        290.00    300.00    310.00    320.00    330.00    340.00    350.00

 1:                                                                        :1
 2:                                            |-- ||||    +||             :2
 3:                                                             -          :3
 4:                              |||                    ||-+               :4
 5:                          ||             ||  ||||-- ||||                :5
 6:                                  ||||                    +||   || |||| :6
 7:                                                       o                :7
 8:                            ||                                          :8
 9:                 |||||||||   +                             --      --   :9
10:                                                                        :10
11:                                                       +|    ||  ||| -- :11
12: |||||||||||||                                                          :12
20:                                                                        :20

             !         !         !         !         !         !         !
        360.00    370.00    380.00    390.00    400.00    410.00    420.00

 1:                             -                     ||||||||  ----       :1
 2:                                                                        :2
 3: - -+||         |||      ||    ||||                                     :3
 4:              ||                    ||--                                :4
 5:             +|        o              |                    ||-||||||  | :5
 6: |||   ||     o       oo--    --       ||||    ||||                     :6
 7:                               o                                        :7
 8:        oo        +| |+-   ||||  -+                                     :8
 9:                   ||          -       o                                :9
10:    o   --                    ||                            +||      || :10
11:   ||    |||||-       ||||                                              :11
12:                                                                        :12
20:                                                                        :20

             !         !         !         !         !         !         !
        430.00    440.00    450.00    460.00    470.00    480.00    490.00
```

```
 1:                                   --                                                        :1
 2:                                   ||||     ||||  ||||||                    -- --            :2
 3:                                                                                             :3
 4:                                                                                             :4
 5: ||          o      ||    ||                --                      --||||oo          - |    :5
 6:                                              |||       +| |||||     || -+|||||||||          :6
 7:                                                                                             :7
 8:  +|||||||+|||||-+|||-||                                                 o                   :8
 9:                        |||        ||                                                        :9
10:          -+            --   |||||                       +|  --+|                            :10
11:                                                                    +|||                     :11
12:                                                                                             :12
20:                                                                                             :20
            !         !         !         !         !         !         !
         500.00    510.00    520.00    530.00    540.00    550.00    560.00
 1:                                                    ||||                                     :1
 2:                                                                                             :2
 3:           ||                                                                                :3
 4:                                   ||            ||||  +||                             |||   :4
 5: ||||| ||||           |||||o|||                                                 ---          :5
 6:     --                                                                       -+|||||||      :6
 7:                                                                                             :7
 8:     |        ||||||||    ||      |||||                                                      :8
 9:                                                                                             :9
10:            +||           --                  --                                             :10
11:                 --                                       |||||||||||||||                    :11
12:                                                                                             :12
20:                                                                                             :20
            !         !         !         !         !         !         !
         570.00    580.00    590.00    600.00    610.00    620.00    630.00
653.02  wrap at addr 0,33354 (oct) in file
 1:                                                                                             :1
 2:                                                      ||||||||                               :2
 3:                                                                                             :3
 4: |  +|                                                                                   +   :4
 5:          ||        ||                                      ||||||||||||||   ||||            :5
 6: -+||       ||      ||        || || ||                                                       :6
 7:                                                                                             :7
 8:  -                                   ||  -+|                                 ||||  --       :8
 9:                                                                                             :9
10: ||-+ |||||    --+||                                                                         :10
11:               |||                                                                           :11
12:                                                                                             :12
20:                                                                                             :20
            !         !         !         !         !         !         !
         640.00    650.00    660.00    670.00    680.00    690.00    700.00
```

```
 722.00  end at addr 0,35510 (oct) in file
1:                                                                          :1
2:                                                                          :2
3:                                                                          :3
4: ||                                                                       :4
5:     -- +||||||||||||||                                                   :5
6:                                                                          :6
7:                                                                          :7
8:         --                                                               :8
9:    |||||                                                                 :9
10:  |||                                                                    :10
11:                                                                         :11
12:                                                                         :12
20:                                                                         :20
         !        !        !        !        !        !        !
     710.00   720.00   730.00   740.00   750.00   760.00   770.00
```

```
last time stamp at    722.00
  3730 time stamp words in file
maxdur =    584.88 seconds at addr 35510 (octal)
     0 marker words in file
```

Counts of changes by phones into these states

pho	0	1	2	3	4	5	6	7	total
1	30	2	62	62	8	8	0	1	173
2	37	8	132	131	22	23	0	1	354
3	20	4	42	42	6	8	0	0	122
4	31	6	102	102	13	18	0	3	275
5	71	13	211	209	15	16	0	1	536
6	68	14	330	330	27	35	0	2	806
7	20	7	27	27	14	16	0	0	111
8	70	14	126	125	19	22	0	3	379
9	38	8	82	81	15	17	0	2	243
10	40	9	88	87	25	29	0	5	283
11	40	8	171	171	21	24	0	1	436
12	10	1	62	62	0	0	0	0	135
20	1	1	0	0	0	0	0	0	2
tot	476	95	1435	1429	185	216	0	19	3855

Counts of durations in various states

dur (ms)	0	1	2	3	4	5	6	7
0- 0	0	0	1	0	0	0	0	0
20- 20	4	36	149	135	23	45	0	19
40- 40	11	11	114	160	10	30	0	0
60- 60	14	8	135	176	15	26	0	0
80- 80	13	6	104	155	12	23	0	0
100- 100	10	6	77	150	20	13	0	0
120- 120	8	7	79	109	6	17	0	0
140- 140	10	3	69	98	3	14	0	0
160- 160	7	5	45	76	3	10	0	0
180- 180	3	4	43	75	2	9	0	0
200- 200	4	2	44	57	5	8	0	0
220- 220	9	1	35	53	5	1	0	0
240- 240	6	0	31	44	1	7	0	0
260- 260	2	1	20	37	4	4	0	0
280- 280	5	0	19	18	4	3	0	0
300- 300	3	1	16	18	1	0	0	0
320- 320	7	3	21	16	1	1	0	0
340- 340	1	0	8	12	0	0	0	0
360- 360	4	0	17	10	0	2	0	0
380- 380	3	1	8	7	0	1	0	0
400- 400	4	0	12	5	1	1	0	0
420- 420	5	0	8	4	1	0	0	0
440- 440	1	0	10	5	0	0	0	0
460- 460	3	0	4	0	0	0	0	0
480- 480	3	0	6	3	1	1	0	0
500- 500	2	0	93	1	44	0	0	0
520- 520	2	0	267	1	0	0	0	0
540- 540	9	0	0	1	1	0	0	0
560- 560	3	0	0	0	3	0	0	0
580- 580	0	0	0	0	2	0	0	0
600- 600	2	0	0	1	0	0	0	0
620- 620	2	0	0	0	2	0	0	0
640- 640	3	0	0	0	2	0	0	0
660- 660	0	0	0	0	2	0	0	0
680- 680	3	0	0	0	1	0	0	0
700- 700	3	0	0	0	0	0	0	0
720- 720	4	0	0	1	0	0	0	0
740- 740	5	0	0	0	1	0	0	0
760- 760	1	0	0	0	1	0	0	0
780- 780	1	0	0	0	0	0	0	0
800- 800	1	0	0	0	0	0	0	0
820- 820	2	0	0	0	1	0	0	0
840- 840	2	0	0	0	0	0	0	0
860- 860	0	0	0	0	0	0	0	0
880- 880	1	0	0	0	0	0	0	0
900+...	303	0	0	1	7	0	0	0

Number of seconds each phone is in each state

pho	0	1	2	3	4	5	6	7
1	693.22	0.06	18.84	5.58	3.94	0.34	0.00	0.02
2	672.02	0.82	27.14	14.40	5.26	2.34	0.00	0.02
3	701.00	0.18	13.20	4.86	1.60	1.16	0.00	0.00
4	677.74	0.34	22.80	16.22	3.06	1.78	0.00	0.06
5	637.46	1.32	46.26	29.32	6.34	1.28	0.00	0.02
6	606.06	1.04	57.80	47.32	5.84	3.90	0.00	0.04
7	705.30	0.66	8.58	3.22	2.68	1.56	0.00	0.00
8	661.58	1.06	37.22	12.52	7.06	2.50	0.00	0.06
9	680.50	0.62	26.06	7.38	5.78	1.62	0.00	0.04
10	678.34	0.84	21.24	10.42	8.04	3.02	0.00	0.10
11	659.90	0.80	31.86	22.80	4.42	2.20	0.00	0.02
12	700.26	0.02	15.22	6.50	0.00	0.00	0.00	0.00
20	721.74	0.26	0.00	0.00	0.00	0.00	0.00	0.00

Total times: n phones simultaneously > threshold (bit0=1)

#ph	time (seconds)
0	521.96
1	190.82
2	8.00
3	1.06
4	0.16
5+	0.00
sum	210.64

Total time spent by each speaker speaking (spurts)

spkr	time (secs)	(each x = 2 secs, rounded up)
1	8.22	xxxx
2	28.86	xxxxxxxxxxxxxx
3	8.54	xxxx
4	27.02	xxxxxxxxxxxxxx
5	46.94	xxxxxxxxxxxxxxxxxxxxxxx
6	79.32	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
7	6.68	xxx
8	19.30	xxxxxxxxxx
9	13.56	xxxxxxx
10	21.50	xxxxxxxxxxx
11	38.16	xxxxxxxxxxxxxxxxxxx
12	11.98	xxxxxx
20	0.26	

# of times speaker had talk spurts this many seconds

spkr	0.00-0.38	0.40-0.78	0.80-1.18	1.20-1.58	1.60-1.98	2.00-2.38	2.40-2.78	2.80-3.18	3.20-3.58	3.60-3.98	4.00-
1	9	5	3	0	0	0	0	0	0	0	0
2	9	7	8	4	1	1	1	1	0	0	0
3	3	8	2	1	0	0	0	0	0	0	0
4	6	7	8	3	2	0	2	0	0	0	0
5	10	25	8	7	4	1	0	1	0	0	0
6	9	19	12	5	4	4	1	1	3	0	3
7	4	4	1	0	0	1	0	0	0	0	0
8	11	11	3	3	0	0	0	1	0	0	0
9	9	8	5	1	0	0	0	0	0	0	0
10	7	7	6	2	2	0	0	1	0	0	0
11	5	12	8	1	1	3	1	2	0	1	0
12	3	6	1	2	1	0	1	0	0	0	0
20	1	0	0	0	0	0	0	0	0	0	0

number of talk spurts = 344
sum of all talk spurts = 310.34 secs
mean spurt size = 0.90 secs

Exhibit D.5.1 Sample Audit Trail

APPENDIX E

CONFERENCING FACILITY

APPENDIX E
CONFERENCING FACILITY

1.0 CONFERENCING HARDWARE

Basically, the system consists of two independent sections – a control section and an audio conditioning section. The control section is composed of 20 touch-tone data sets connected to dial-up Bell System lines. These lines are automatically answered to establish a user-to-computer connection, and are then used to transmit touch-tone commands from a user to a PDP 11/45. These commands control conference configurations and conference queues in real time. The audio conditioning section consists of a multiplexed A/D-D/A system and a large buffer memory connected to a signal processing machine (LDVT) which allows audio connections to be made arbitrarily between users. In addition, three ports on the A/D-D/A system are available to connect external voice equipment. The large buffer memory can implement delays of up to 0.5 sec for each of the 20 dial-up users. For additional flexibility, the signal processor is also connected to the 11/45 so that the control inputs can be used to modify the switching and signal processing operations in real time.

1.1 System Description

Figure E-1 is a block diagram of the complete conferencing facility. From the point of view of the PDP 11/45 machine, two external devices are connected through standard DEC interface

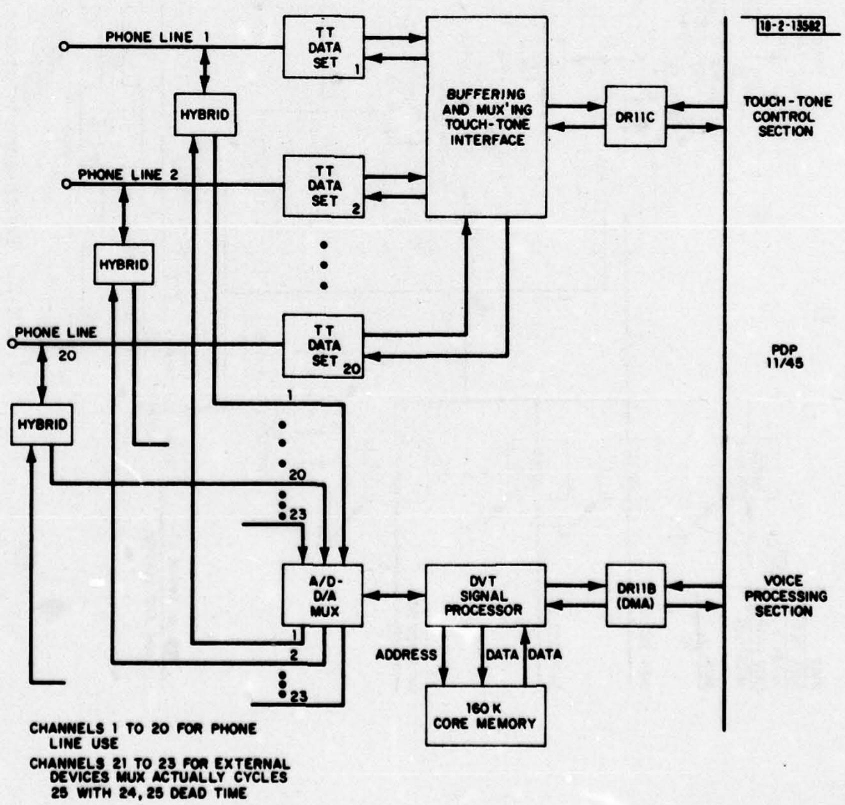


Fig. E-1. Conferencing system.

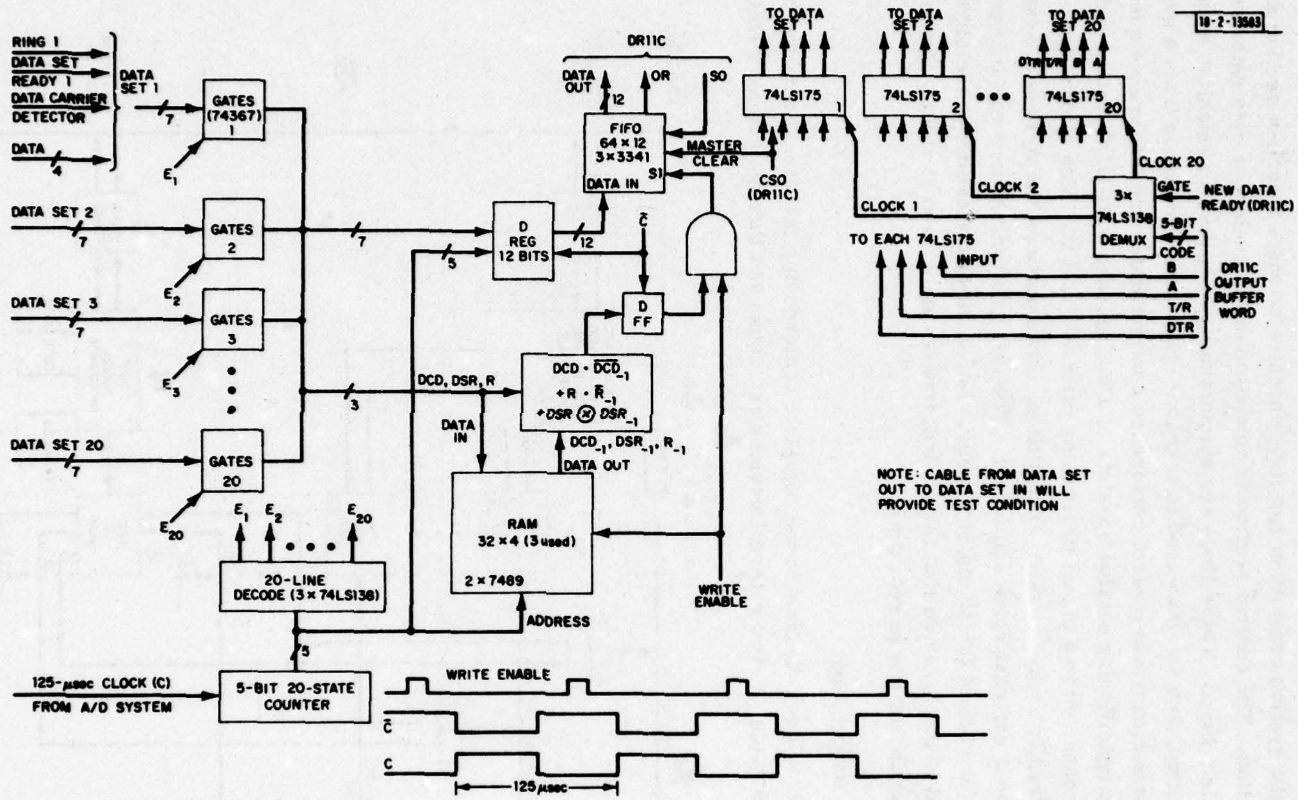


Fig. E-2. Touch-tone interface, 20 data sets to DR11C.

circuits. The telephone control system is connected through a standard DR11C single-word interchange board with interrupt capability. The audio-switching section is connected through a more flexible DR11B direct-memory-access (DMA) interface. Twenty 2-wire phone lines are connected to the touch-tone receivers for the control path, and to a set of hybrid (2- to 4-wire) transformers for the audio path. Four wires from each of the 20 lines are connected to an A/D-D/A converter port for audio switching.

1.2 The Touch-Tone Receiver Control Path

Each of the 20 phone lines is connected to a Bell type 403 tone data set which automatically responds to a ringing signal by passing a ringing bit (R) to the computer interface. If the computer raises a data terminal ready (DTR) bit, the data set will answer the line and set up to receive control tones by transmitting a data set ready (DSR) bit. When a user presses a tone button, the data set will signal the computer with a data carrier detector (DCD) bit, and a 4-bit tone code. The computer can listen for these tones, have the data set transmit three single-frequency responses, or hang up.

Figure E-2 presents the interface between 20 data sets and the DR11C. The basic interface function scans the 20 data sets for activity by comparing a new status word from each channel with a previous stored status word from the same channel. Each previous channel status word has been stored in the 32- × 4-bit RAM. Only the three status bits (R, DCD, and DSR) need be stored for comparison against the latest word. If there is a change in any of these bits where change is defined as: $DCD \cdot \overline{DCD}_{-1} + R \cdot \overline{R}_{-1} + DSR \otimes DSR_{-1}$, then the present word, including a 5-bit code for channel identification, is clocked into a first-in/first-out (FIFO) buffer and an output request is set. The 20 data sets are scanned in a cycle of 20 of the 8-kHz (125-µsec) samples (see Fig. E-3), so that a complete scan requires 20 × 125 × µsec = 2.5 msec. Each data set is controlled from the interface by a 4-bit register which is loaded under program control from the PDP 11/45 – DR11C path.

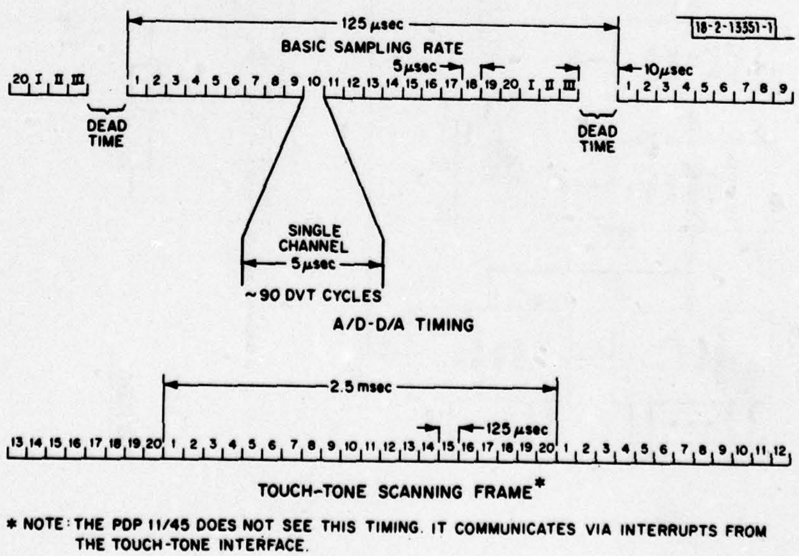


Fig. E-3. Conferencing system timing.

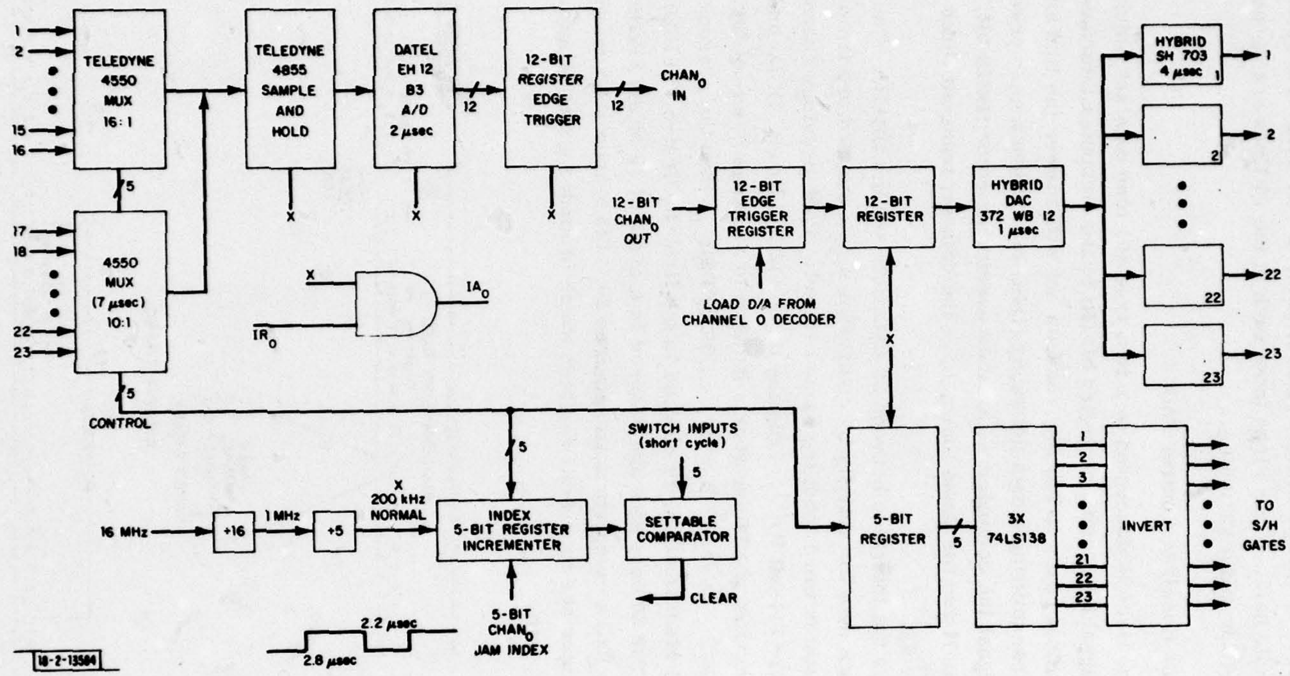


Fig. E-4. A/D-D/A-MUX system.

1.3 The Audio Conditioning Section

As Fig. E-1 indicates, the audio conditioning section consists of three subsections: an LDVT signal processing computer, a multiplexed A/D-D/A system which is controlled by and communicates with the LDVT, and finally a large (160K) core memory which is controlled by the LDVT. The LDVT, in turn, can also communicate with the PDP 11/45 through a DMA interface called a DR11B.

1.4 The Multiplexed A/D-D/A System

The A/D-D/A system is shown in Fig. E-4. It is connected to the channel 0 input and output ports of the LDVT and consists of an A/D section, a D/A section, and some multiplexing timing registers.

The A/D section can accept up to 32 input analog signals multiplexed through two Teledyne 16:1 gates (only 23 inputs are used). These multiplexer gates drive a sample-and-hold (S/H) gate which drives, in turn, a 12-bit A/D converter. The multiplexed input is controlled from a 5-bit register incrementer which can be loaded with a 5-bit word asynchronously so that random access conversion of any input channel can take place; or, a standard input clock will increment the register by one during each cycle and clear at some settable value. In other words, the input multiplexer can be stepped randomly, or cycled through a fixed pattern. A normal input conversion rate is 200 ksamp/sec, although an external clock can be used. The input A/D 12-bit word is read on input channel 0 of the LDVT, either as a forced input or an interrupt.

The D/A section is double buffered, which means that the user can load the D/A buffer on a channel 0 output from the LDVT but the transition of the D/A converter will take place on the next synchronous clock edge. A demultiplexer S/H gate is controlled by a 5-bit word delayed by one clock cycle from the input MUX control. This allows for the delay in D/A conversion. The D/A section consists of the double buffering, a fast 12-bit D/A converter, a set of 23 (expandable to 32) S/H gates, and a 5-bit decoder pulse steerer. The choice of S/H outputs rather than individual slower D/A registers and converters was based on cost and wiring complexity.

1.5 The Large Buffer Memory and Interface

Basically, the large buffer memory is a 128K by 20-bit core memory plus a 32K by 20-bit core memory, and both have a read-modify-write time of approximately 2 μsec. We have designed a 16-bit word interface, consistent with the LDVT data word length, although our delay experiments will require only 12-bit words. The input to and output from the memory (write and read words) are communicated from and to channel 2 of the LDVT. Actual read, write, read-modify-write, load address, and various hybrid commands to the large memory are transmitted from output channel 0 of the LDVT. Since this channel was designed as a 12-bit output to a D/A converter, 4 more bits are available to be decoded and used to steer data to places other than the D/A converter. The lower-left portion of Fig. E-5, the memory interface and channel 0 decoder, shows the decoding table. An output on channel 0 from the LDVT, with 4 upper bits zero or all 1, produces a standard D/A load. The other commands load upper and lower portions of the 18-bit address register, and start read, write, or read-modify-write cycles. Since the output on channel 0 is a 12-bit word, the loading of the address register is a two-command operation. The lower address (A_L) is 12 bits and the upper portion (A_U) is 6 bits. Presumably, only the lower register would be loaded for many applications requiring only one command. It is also possible to combine the address load with a read, write, or read-modify-write command. Two remaining commands set up the multiplex word and do a master clear.

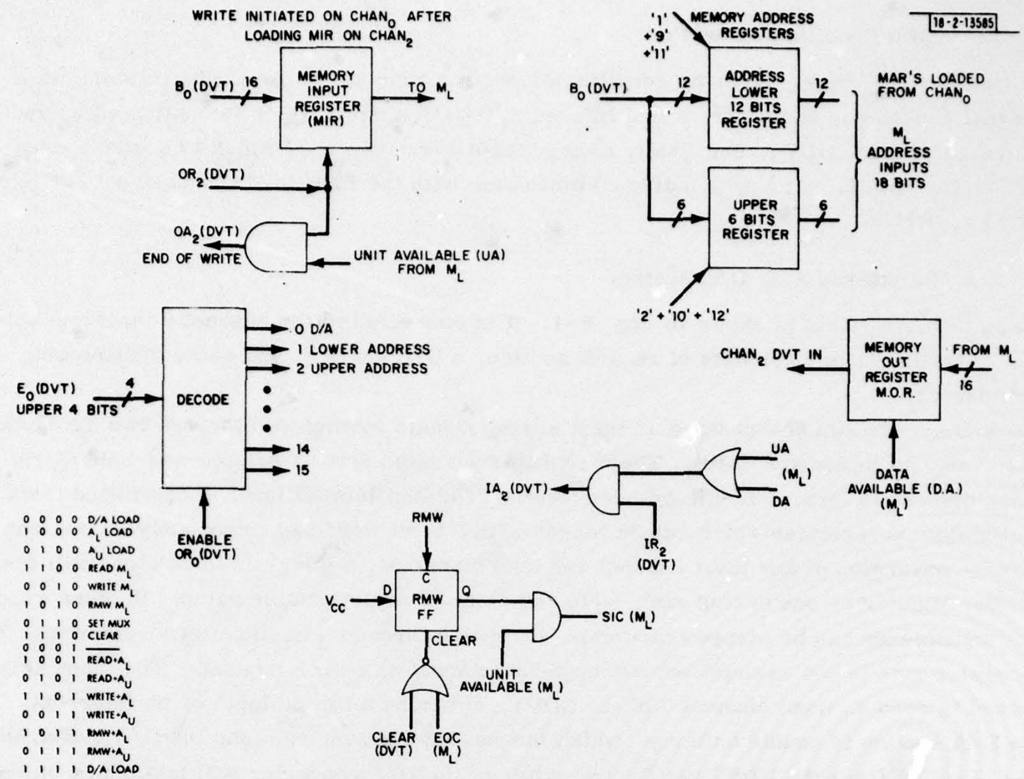


Fig. E-5. Large memory interface (M_L) channel 0 decoder.

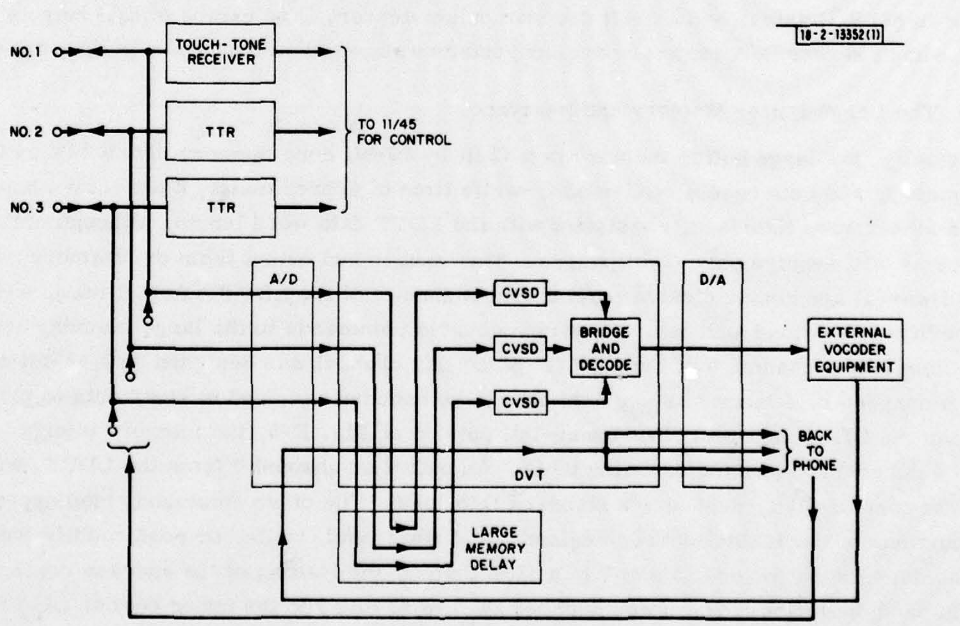


Fig. E-6. Conference example.

1.6 The LDVT as Controller

The LDVT has a limited in-out system which has been modified to control the multiplexing system and the large memory. The present 4 channels of input and output are assigned as follows. Channel 0 outputs to the D/A converter, sets the MUX index, or controls the large memory. Channel 0 input receives data from the A/D converter. Channel 1 communicates with the PDP 11/45 through the DR11B interface. Channel 2 reads from and writes to the large memory (M_L). Finally, channel 3 remains as the link to M_X, the internal LDVT bulk memory. The 55-nsec cycle time of the LDVT allows for approximately 90 machine cycles during each 5-μsec A/D conversion cycle.

2.0 CONFERENCE EXPERIMENT EXAMPLE

Figure E-6 shows the conferencing facility as it might be configured for a 3-party conference. This example shows a conference which is bridged at the delta modulated bit level, tandemed in a narrowband vocoder, and then distributed to the conferees.

The three participants form the conference by dialing up one of the 20 phone numbers, and communicating via touch-tone to the PDP 11/45 conference control program. The LDVT software is loaded via the 11/45 to implement CVSD encoders for each of the participants, effect the bit stream bridging, delay the audio inputs by fixed or time-variable amounts, output the decoded bridged signal to an externally connected vocoder (on channel 21, 22, or 23), and receive the output of the vocoder tandem back on the corresponding A/D channel for distribution to all the conferees, or all except the one talking.

If a fourth person wishes to join the conference, he calls in and interacts with the control software scanning the touch-tone interface. Then, flags are activated in the LDVT to enable another A/D-D/A channel and include the fourth stream in the bridging and distribution.

Figure E-7 indicates the physical layout of conferencing equipment aside from the PDP 11/45, and the large core memory used for delay.

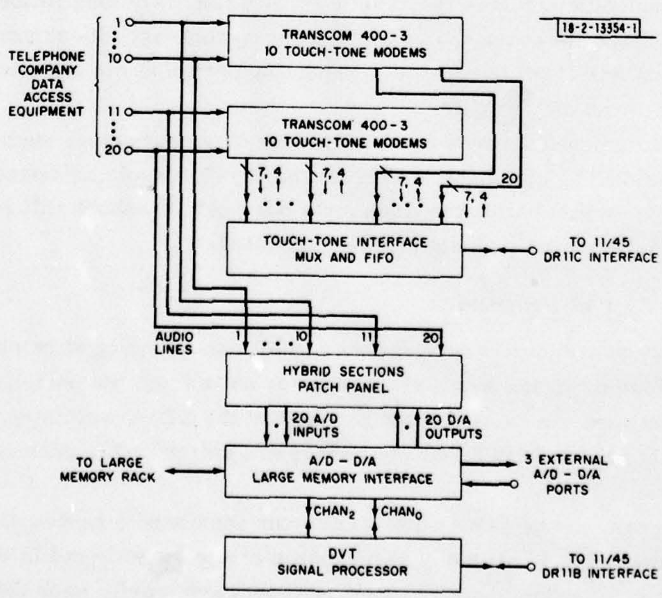


Fig. E-7. Conferencing rack.

As mentioned earlier, statistics about activity, coincidence of talkers, etc. can be gathered on-line by way of the LDVT link to the 11/45.

3.0 CONFERENCING SOFTWARE

In this section, the software which is common to all simulations involving the hardware conferencing facility is discussed. The simulation of a particular conferencing technique is realized by extending this common software base to effect the desired bridging or switching technique. The commonality follows from a decision to fix the information format exchanged between the LDVT signal processor and the PDP-11/45 control and data collection processor. As a result, all voice energy switched and bridged conferencing simulations can use the same 11/45 programs for control, data collection, and data reduction. However, the 11/45 code must be specialized for the conference technique which uses touch-tone signalling.

3.1 PDP-11/45 – LDVT Communications

Communication between the LDVT and the 11/45 involves the transfer of blocks of twenty 16-bit words every 20 msec. There is one word in each block for each possible conference participant. A bit in each word indicates to the LDVT whether the corresponding phone is to be considered active or not. If a phone is marked as active, the LDVT program will treat the signal from that phone according to the conferencing algorithm in effect. In addition, the program will look for speech activity from that phone by accumulating the sum of the absolute values of the PCM readings for each 2-msec interval. If the sum exceeds a threshold during any of the ten 2-msec intervals in a 20-msec reporting period, the LDVT will indicate that fact to the PDP 11 program by setting a speech activity bit in the corresponding word of the block sent to the 11/45.

If the conferencing technique being simulated involves signal selection based on voice energy detection, the LDVT program carries out the decision logic and indicates its decision by setting another bit in the communication word corresponding to the selected speaker. If a speaker/interrupter technique is being simulated, yet another bit is set to mark the interrupter. The 20-msec reporting period determines the resolution at which switching times are known, but the actual switching instant is quantized by the 2-msec speech-activity accumulation time. The loss in resolution resulting from the 20-msec reporting period is not significant since speaker switching occurs at a much slower rate.

The block of communication words can be used for other purposes, such as allowing timing and amplitude threshold to be communicated from the 11/45 console keyboard to the LDVT which lacks console control. In simulations to date, only one such threshold value is used. Its meaning varies with the conferencing technique being simulated.

3.2 PDP-11/45 Control Program

The 11/45 control program has two functions in all conferencing simulations. One is to command the touch-tone interface hardware to answer calls from the participants and thus effect the connection between the participants' phones and the LDVT switching/bridging processor. The other is to indicate to the LDVT that the phones are active and to pass run-time parameters to the LDVT program.

The control program can be given commands from the console keyboard to indicate which phones are to be answered and how many participants are to be accepted in the conference. While it is possible for n active phones to be distributed arbitrarily over the available phone

numbers, current software limits a conference of n participants to the first n phones in the order of their connection to the conferencing hardware.

Two versions of the control program are available. In the first (the most commonly used), the commands to answer the first n phones are issued prior to any participant dialing activity. In the second, the control program waits for a ringing signal and issues the command to answer when the ringing signal is observed and the phone is one of the n to be accepted. In the first case, all conference participants hear the tone generated by the answering hardware and are made aware that someone is entering the conference. In the second case, the tone is inhibited because the control program does not tell the LDVT that the new phone is active until the end of the answering tone.

3.3 LDVT Switching/Bridging Program

The LDVT program receives PCM inputs and provides outputs for all phones connected to the conferencing facility. The A/D-D/A multiplexer scans through the 20 phone lines and three speech encoder ports, allowing 5 μsec per line for processing in the LDVT. This time allows approximately 90 instructions to be executed in the LDVT for each phone line. These instructions must provide for the execution of the basic signal selection or bridging algorithm required for the conferencing technique being simulated, as well as to allow for speech-activity detection and, in some cases, delays corresponding to satellite transmissions. In addition, small delays are introduced to improve the operation of the speech-activity detectors in voice-switched signal selection conferences by allowing the detector to anticipate threshold crossings. The delays are realized by storing the PCM speech samples in a large core memory attached to the LDVT. When satellite delays as well as anticipatory delays are used, almost all of the possible 90 instruction executions are needed. Very careful coding is required to avoid exceeding the 5-μsec timing constraint.

The exchange of information with the 11/45 is handled in the 10 μsec which remain in the basic 125-μsec frame (8-kHz sampling rate) after servicing the 20 phone lines and three speech encoder ports. Word transfers take place on 40 (20 in each direction) of the 160 frames which occur during the 20-msec reporting period. The transfers are spaced to allow the slower 11/45 hardware and software to handle them without difficulty.

3.4 PDP-11/45 Data Collection Program

As discussed above, the LDVT sends 3 bits of information to the PDP-11/45 for each participant during every 20-msec reporting period. These bits tell whether or not the participant was exhibiting speech activity, was the selected speaker, or was the selected interrupter during the previous period. We call the combination of these 3 bits the "state" of the participant. The data collection program observes the state of each participant and makes up a disk file which has an entry for every change of state. The entry shows the new state, as well as a time marker equal to the number of 20-msec report periods since the start of the conference.

Data collection begins when the conferencing simulation program starts, and ends when the program is manually stopped. To allow experimenters to mark off time periods of interest during a conference, a push button is available which when pushed introduces a signal into an otherwise unused phone channel. The signal is noted in the collected data. A companion button can be pushed to add an audible tone to the audio recording normally made during a conferencing experiment. This tone can be used to correlate the marked point in the data with the conference content.

3.5 Data Reduction Software

To aid in the analysis of conferencing experiments, a data reduction program has been developed which produces both global summary information and/or a detailed step-by-step history of the conference interactions. The data reduction program operates on the files established by the data collection program.

The global summary information is produced in the form of several charts:

(1) For each speaker, a count of the number of times a transition was made into each of the possible states.

(2) For all the speakers combined, a histogram of the durations in the various states.

(3) For each speaker, the total time spent in each state.

(4) For each speaker, the total time spent speaking, i.e., the sum of the speaker's talkspurts. A talkspurt is defined as a "smoothed" time interval when the speaker's energy level was above threshold. To provide the smoothing, "small" silence gaps (i.e., intervals in which energy is below threshold) are considered as part of the talkspurts. After these "silence gaps" are filled, any resulting talkspurts that are suitably small are considered irrelevant noise and are disregarded in the final tabulation. The two constants, the size of the silent gaps and the size of the ignored spurts, are easily modified. This method of tabulating talkspurts closely simulates the perceptions by humans who normally consider a talkspurt as a substantial interval between major silences, ignoring small silences between syllables or words.

(5) For each speaker, a histogram of the number of talkspurts of various durations. Talkspurts are as defined in the previous paragraph.

(6) For the conference as a whole, the total times n phones were simultaneously over threshold. This provides a convenient measure of the amount of talk as well as conflict (simultaneous talk) in the conference. It should be noted that the feature measured here is simply energy level above threshold rather than talkspurts. An example of summary outputs from a conference experiment is shown in Appendix D.

A detailed step-by-step picture of the conference is provided by an audit trail output. The time axis extends horizontally and speakers are plotted in the vertical axis, analogous to a strip-chart recording. At each intersection of time and speaker, an indication of the state of that speaker for that time interval is presented.

Duration of time interval is selectable when the audit trail program is run. Two distinct audit trails are available based on the selection of time interval for each tick mark in the time axis.

If each tick mark is selected to be one 20-msec period, then the audit trail shows each actual transition as it occurs. No merging need be done, since 20 msec is the basic time unit for indicating transitions to the 11/45.

If the tick mark is more than one 20-msec period, then the audit trail shows merged information. For example, during a 1-sec interval (fifty 20-msec periods) a given speaker may have been both the designated interrupter and the designated speaker. The example in Appendix D is an audit trail with a 1-sec marking interval.

In addition to providing global summary information and step-by-step pictures, the analysis is useful as an aid to debugging and fine-tuning the conferencing algorithms.

4.0 DISCUSSION OF FACILITY LIMITATIONS

The use of the telephone system as a part of the conferencing facility has placed some limitations on the performance and capability of the simulation facility. Noise and distortion in the system itself, as well as in the Data Coupler used to interface the conferencing gear to the phone system, result in overall speech quality which is less good than would be expected in a digital communication system with handsets directly connected to speech encoders. In addition, the hybrid transformers which convert the 2-wire phone lines to 4 wires for connection to the A/D-D/A equipment introduce some artifacts which would not be found in a true 4-wire system. Because the hybrid cannot be balanced exactly, some of the signal sent out to each phone line returns as input. The level of this reflected signal is substantially lower than a normal input, and it poses relatively little problem in signal selection conferencing, but in a summation conference with delay it results in a speaker hearing an echo of his or her own voice with a delay equal to twice the simulated communication delay. The magnitude of this reflected signal was made almost independent of the number of conferees by alternating the polarities of the input connections so that the reflection from one phone line would tend to cancel that from the next. The resulting overall echo amplitude was about 30 dB below normal listening level. Such a level of echo is clearly audible and intelligible if listened to intently, but it can be ignored relatively easily and does not interfere with a person's ability to speak as does a high-level delayed echo.

The hybrid reflection could cause a problem with voice-controlled signal selection conferences only if a participant spoke very loudly causing the reflected signal to exceed the speech activity threshold. While we could observe this effect during equipment checkout tests, it was not a problem during conferencing experiments because the subjects did not have occasion to speak so loudly. However, since the reflection added to any real noise present at the input, it forced us to use a higher threshold than would be needed in a true 4-wire system.

A further consequence of the use of the phone system was our inability to experiment with the use of switched sidetone as a means of signalling to a participant that he or she is the selected speaker. In the phone system the sidetone is inherent in the 2-wire connection and cannot be shut off.

APPENDIX F

CVSD MAJORITY VOTING BRIDGE

APPENDIX F

CVSD MAJORITY VOTING BRIDGE

For delta modulation encoding techniques, it is possible to approximate the action of an analog bridge without decoding the signals to be summed. For example, in CVSD encoding, 2-bit sequences may be interpreted as follows:

00	slope is consistently negative
01	slope changes from negative to positive
10	slope changes from positive to negative
11	slope is consistently positive

In our majority voting bridge, the most recent 2 bits from each input encoder are examined and votes are indicated as follows:

00	cast a vote for a negative output slope
01, 10	cast no vote (abstain)
11	cast a vote for a positive output slope

If the majority of input encoders indicate votes for a negative output slope, an output of "0" is generated. If the majority vote is positive, an output of "1" is generated. If a tie vote is registered or all inputs abstain, then the output is set to the complement of the previous output.

The output of such a majority voting bridge exhibits a signal-to-noise ratio (SNR) which becomes progressively worse as the number of inputs increases. The noise increases because the voting process gives equal weight to all input slope information without regard to the magnitude of such changes. We feel that the noise increase limits use of the technique in its pure form to small conferences with, at most, three or four participants.

In order to increase the utility of the majority voting technique and extend it to larger conferences, we have added speech-activity detection to the bridge so that only those phone lines on which activity is detected are considered in the voting procedure. As a result, since most of the time in a conference only one participant is speaking, the speech quality will most of the time be no worse than one would expect from CVSD encoding. Only when two or more people speak at the same time (the order of 5 percent of the total speech time in our experiments) is there any degradation of the SNR due to the majority voting operation.

In our implementation the CVSD analysis, the majority voting, and output synthesis are all handled by the LDVT switching/bridging processor. Because of the heavy computing load associated with the CVSD analysis of the input signals, the simulation is limited to eight participants and the transmission delay option is not available.

In order to achieve 16-kbps CVSD speech encoding with the conferencing A/D multiplexer which runs at an 8-kHz rate, it is necessary to estimate every other sample by means of linear interpolation. This technique introduces a negligible error when the input speech is band-limited correctly for the 8-kHz sampling rate, as it should be for 16-kbps CVSD encoding.

Unlike the analog bridge simulation, the CVSD majority voting bridge does not subtract out a speaker's voice from the signal he or she hears, because the LDVT cannot handle the computations required to produce eight different outputs. Since this simulation does not include delay effects, the speaker hears this as normal sidetone.

APPENDIX G

CONTROL SIGNAL SELECTION (CSS) SYSTEM

APPENDIX G
CONTROL SIGNAL SELECTION (CSS) SYSTEM

The system that we implemented to explore control signal selection techniques made use of the tone keys normally used for dialing in modern telephone systems. Participants pushed keys to indicate to the controller their desire to speak and the fact that they were finished speaking. The controller signaled to the participants by sending combinations of three tones in a variety of time patterns. The tones were generated by the Touch Tone data set hardware on signal from the PDP-11/45 control computer.

There are very many possibilities for the use of the buttons and signaling tones in a CSS conferencing system. We tried a number of variations before settling on the particular choices described in this appendix and used in the Phase II experiments. In particular, we tried several different sets of signals to find a set in which all the signals were readily identifiable and easy to remember. It is easy to provide an overlay to remind users of the functions of the keys, but it is harder to provide useful aids for interpreting the signals.

Participants other than the chairperson had only two active keys. The "want-to-talk" (WTT) key was key "0," the middle key in the bottom row. Pushing WTT while someone else was speaking caused the controller to send an acknowledgment signal and to place the participant in the queue of participants waiting for a chance to talk. The acknowledgment signal was an ascending series of the three available tones each held for 160 msec and generated without intervening pauses. The effect was that of a single burst of shifting pitch which we called a "bleep." The second key was used to indicate that the speaker was finished talking or that he or she wished to be removed from the queue. This "done talking" key was the "*" key on the lower left corner of the key pad. The controller signaled receipt of the DONE key by sending an alternating sequence of two tones which produced a warbling effect. The period of the warble was 100 msec, and the signal lasted for 400 msec. The acknowledging bleeps and the warbles were heard only by the participants who pushed the keys which elicited the responses.

When the controller received a WTT signal from a participant, it placed him or her on a queue of persons wanting to talk. Barring special actions by the chairperson, the queue was processed on a first-come, first-served basis. When a participant was about to be given the floor, the controller sent a "you are now on the air" signal consisting of three short tone bursts followed by a longer burst at a lower pitch. The signal was similar in its aural effect to the opening motif of Beethoven's Fifth Symphony. It lasted for 780 msec. The actual selection of the participant as speaker took place at the end of the signal so that other participants would not hear the signal.

Once given the "floor," a speaker was allowed to talk until either he or she had pushed the DONE key, a timer ran out, or the chairperson intervened. The timer ran only when some other participant was in the queue, and in our experiments, was set to a relatively long time (40 sec) so that it rarely operated to cut off a speaker. The timer did not run while the chairperson held the floor. When a speaker was about to be timed out, the controller sent a warning signal consisting of four 140-msec tone bursts separated by silent intervals of the same duration. This warning signal was heard by the other participants as well as the speaker because it was introduced on the 2-wire side of the hybrid transformer that connected the speaker's phone line to the conference controller. The subjects indicated that they felt it was useful to hear the speaker being warned. If the speaker did not finish talking and push the DONE key within 7 sec

after the warning signal, the controller would switch to the next speaker in the queue. In this case, the speaker being cut off would hear the same warble signal associated with pushing the DONE key.

The chairperson had four additional active keys which he or she could use to effect some control over a conference; these were:

(1) Interrupt the Current Speaker (Key "9"). Pushing this key would preempt the floor from the current speaker and give it to the chairperson. The preempted speaker would be placed at the head of the queue so that he or she would automatically resume as speaker when the chairperson pushed DONE.

(2) Priority Want-to-Talk (Key "8"). Pushing this key would put the chairperson at the head of the queue so that he or she would become the next speaker when the current speaker finished.

(3) Force Timeout of Current Speaker (Key "7"). Pushing this key would cause the current speaker to be given the warning signal and then to be cut off within 7 sec if he or she did not voluntarily relinquish the floor by pushing the DONE key.

(4) Axe the Queue (Key "4"). Pushing this key would cause the controller to forget all queued requests to speak. Its use was appropriate in situations where the chairperson wished to change the topic of discussion. The first action would be to seize the floor and announce the desired change. Since the queue (if any) held people who presumably wanted to talk about the old topic, there would be little reason to suppose that they would also be ready to talk about the new topic. Axing the queue could avoid people being given the floor only to say that they had nothing to say.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ESD-TR-79-78	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Voice Conferencing Technology Program		5. TYPE OF REPORT & PERIOD COVERED Final Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) James W. Forgie P. L. Weene C. E. Feehrer		8. CONTRACT OR GRANT NUMBER(s) F19628-78-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 33216K
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Communications Agency 8th Street & So. Courthouse Road Arlington, VA 22204		12. REPORT DATE 31 March 1979
		13. NUMBER OF PAGES 202
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

voice conferencing technology
signal summation technique
wideband PCM signals
conferencing protocol
human factors
tandem encoding
collisions
voice control
signal selection techniques
narrowband encoding techniques
speech quality

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report has been written at the end of two years of research on voice conferencing technology. The goal of the research has been to recommend and demonstrate the best secure voice conferencing techniques for future defense communication needs. The focus of the work has been on the human factors aspects of conferencing, an area in which little research had been carried out prior to the initiation of this effort. The report has been prepared as a joint effort by Lincoln Laboratory and Bolt Beranek and Newman Inc., who have carried out the human factors aspects of the research under contract with Lincoln Laboratory.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)